

Jailson Valdique Semedo Lopes

**Distribuição de segmentos do cromossoma X com
ancestralidade europeia e africana na população de Cabo
Verde: implicações para o estudo da miscigenação em
populações humanas e para a história do povoamento do
arquipélago**



Porto

2011

Tese de Mestrado em Biodiversidade,
Genética e Evolução apresentada à
Faculdade de Ciências da Universidade
do Porto

Agradecimentos

Ao Professor Doutor Jorge Macedo Rocha, meu orientador, por me ter acompanhado de perto durante o meu percurso académico e sem o qual este trabalho não seria realidade. Gostaria de lhe agradecer também por ser sobretudo meu tutor e um dos melhores amigos durante a minha estadia em Portugal. Muito obrigado!

À Doutora Sandra Beleza pela ajuda que me deu e me ofereceu sempre que precisasse e pela forma amigável que me tratou durante todos esses tempos;

À Joana Campos quem me acompanhou com muita amizade durante o desenvolvimento de todos os trabalhos de base do laboratório;

À Doutora Susana Seixas, à Isabel e à Margarida por me terem ajudado com muita paciência nos trabalhos de laboratório e análise de dados;

À Patrícia pelo apoio e companheirismo durante todo o tempo que passei no laboratório;

Ao meu amigo Venceslau pela força que me tem dado durante esse percurso;

À minha colega Dora Henriques;

Ao meu amigo João Pedro. Um abraço;

À família Rocha por me ter acolhido como se fosse um elemento da família. Meus sinceros agradecimentos!

Aos meus amigos cabo-verdianos mais chegados no porto com quem “matei saudades” de Cabo Verde e passei muitos momentos de reflexão e convívio. Cito: Carlitos (tcharles), Avelino (Vy), Décio(Béto), Èdi, Frei Elvino Reis (nosso conselheiro), e Budi;

Ao Senhor Martinho, Presidente da associação dos cabo-verdianos no Norte de Portugal;

Ao meu senhorio Pedro Silva pelo alojamento e pela companhia;

Ao grupo de “futebolada” liderado pelo Senhor Mário, pelos jogos das quartas-feiras, pela amizade e pelo convívio;

Às minhas famílias residentes em Portugal e na Fança nomeadamente Mafalda, Taly, Linete, Susana, José Pedro, Custódio, Sulécia, Euclides, Cuia, Zá, Dá, Bela, Raul e António Pedro pela força;

A todos os professores do curso e principalmente aos Professores Doutor Nuno Ferrand e Doutor Alexandrino pela força e pelo apoio académico que me deram durante todos esses tempos;

Aos meus colegas do mestrado e particularmente ao Duarte, Joel, Nuno, Ricardo, Manuel e Miguel pela companhia nas viagens ao Vairõa, pelas noites passadas na Faculdade a trabalhar, pelo apoio, pela amizade e pela convivência;

À minha mãe Adelaide pelo amor e carinho e pelo apoio incondicional à minha escolha profissional. Muito obrigado!

A toda minha família que mesmo longe me apoiaram nessa caminhada;

À Dra. Isabel Inês Araújo pela força e amizade durante todos esses tempos;

À Dra. Ana Hoffer Almada;

À Fundação Calouste Gulbenkian por me ter presenteado os dois anos de bolsa, indispensável à realização desse Mestrado;

À Faculdade de Ciências da Universidade do Porto (FCUP);

Ao Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO);

Ao Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP) e ao seu Presidente Professor Doutor Sobrinho Simões;

À Universidade de Cabo Verde (UniCV) e o seu anterior reitor, Doutor António Correia e Silva;

Agradeço a Deus pela força de vontade e coragem;

Agradeço a todos que contribuíram, directa ou indirectamente, na realização desse trabalho.

Índice

Resumo	1
Summary	2
1. Introdução	3
1.1 História do povoamento de Cabo Verde	3
1.2 Miscigenação e ancestralidade de fragmentos cromossômicos	5
2. Material e Métodos	9
2.1 Composição da amostra	9
2.2 Marcadores do cromossoma X	9
2.3 Delimitação de fragmentos cromossômicos com proveniência africana ou europeia	9
2.4 Genotipagem de SNPs e microssatélites	12
3. Resultados e Discussão	13
3.1 Resultados obtidos com SNPs	13
3.1.1 Distribuição de segmentos cromossômicos com ancestralidade africana e europeia	13
3.1.2 Estimativas de miscigenação	16
3.1.2.1 Distribuição da miscigenação individual no arquipélago	16
3.1.2.2 Comparação da miscigenação individual nas diferentes ilhas	17
3.1.2.3 Comparação da miscigenação em diferentes compartimentos do genoma	20

3.1.2.4	Miscigenação e recombinação	22
3.1.2.5	Tempo decorrido desde o início da miscigenação	24
3. 2.	Resultados obtidos com STRs e SNPs	27
4.	Conclusões	31
5.	Referências bibliográficas	33

Resumo

As populações miscigenadas que resultaram do contacto recente de grupos humanos diferenciados são experiências naturais que podem ser usadas para estudar a dinâmica da miscigenação e a biologia de características complexas. Com o objectivo de caracterizar o processo de miscigenação em Cabo Verde estudou-se a distribuição de fragmentos de ancestralidade europeia e africana em cromossomas X de 210 indivíduos, não aparentados, do sexo masculino provenientes das 6 ilhas mais povoadas do arquipélago: Santiago (n=25), Fogo (n=64), Boavista (n=21), Santo Antão (n=42), São Nicolau (n=34), São Vicente (n=24). A identificação da ancestralidade de fragmentos cromossómicos foi feita com 14 microssatélites escolhidos ao acaso e 37 polimorfismos nucleotídicos simples (SNPs) com diferenças de frequências alélicas entre Europa e África >0,60 (densidade média de 1 marcador por cada 3,53 cM). Nas análises realizadas só com os 37 SNPs foi possível imputar ancestralidades alélicas com um nível de ambiguidade relativamente baixo (12,5%) que contrasta com a percentagem de incerteza de 86% obtida quando apenas se usaram microssatélites. Quando se usaram simultaneamente SNPs e microssatélites, a incerteza foi pouco maior (18,2 %) do que a correspondente aos SNPs. Os resultados obtidos só com SNPs não diferiram significativamente dos obtidos com SNPs e STRs. Estimou-se uma proporção média de ancestralidade africana nos cromossomas X de Cabo Verde de 65%, concordante com os dados históricos que indicam que a miscigenação no arquipélago foi assimétrica e envolveu quase exclusivamente uniões entre homens europeus e mulheres africanas. A média das ancestralidades africanas calculadas em cada ilha foi de 0,81 em Santiago, 0,68 em Santo Antão, 0,64 em São Vicente, 0,62 em São Nicolau, 0,61 no Fogo e 0,59 na Boavista. As diferenças na contribuição africana para as várias ilhas está globalmente de acordo com os ciclos históricos de povoamento do arquipélago, embora a população do Fogo tenha níveis de miscigenação maiores do que aparentemente seria de esperar, tendo em conta que aí prevaleceu um sistema escravocrata de características semelhantes às de Santiago. A distribuição da miscigenação nas diferentes ilhas está qualitativamente de acordo com os níveis de miscigenação calculados com o cromossoma Y e com *loci* autossómicos, embora os valores de ancestralidade africana no cromossoma X sejam menores do que seria de esperar de acordo com modelos simples de panmixia. A percentagem de transições de ancestralidade ao longo do cromossoma X está fortemente correlacionada ($R^2=0,86$) com a fracção de recombinação, tendo-se verificado uma queda acentuada da recombinação na região do centrómero, onde foi possível definir um segmento de ~14 Mb que provavelmente retém a ancestralidade original de cada cromossoma. Com base na identificação desta região explorou-se um método de datação do início da miscigenação que gerou estimativas entre ~330 e ~ 500 anos, compatíveis com a documentação histórica. A observação de correlações significativas entre o tamanho de fragmentos cromossómicos de proveniência europeia ou africana e os níveis de miscigenação, em conjunto com a evolução da miscigenação em cromossomas de origem europeia e africana, indicam que o processo de miscigenação não atingiu ainda o equilíbrio. Estes resultados constituem a primeira tentativa de caracterizar a miscigenação em Cabo Verde com o nível analítico de maior complexidade.

Summary

Recently admixed populations are natural experiments that may be used to study both the dynamics of admixture processes and the biology of complex traits with biomedical or anthropological interest. To characterize the admixture process in the population of Cape Verde, we studied the distribution of chromosome X fragments with European and African ancestry in 210 unrelated male individuals from the 6 most populated islands of the archipelago: Santiago (n=25), Fogo (n=64), Boavista (n=21), Santo Antão (n=42), São Nicolau (n=34), São Vicente (n=24). Fragment ancestry was identified with 14 randomly chosen microsatellites and 37 single nucleotide polymorphisms (SNPs) with allele frequency differences >0.6 between Europe and Africa (average marker density of 1 marker each 3.53 cM). Uncertainty in ancestry imputation was relatively low (12.5%) in the analyses that were performed only considering SNPs, in contrast with the high (86%) ambiguity levels obtained when using only microsatellites. When SNPs and microsatellites were used together, the uncertainty was only slightly higher (18.2%) than with SNPs alone. Results obtained with both SNPs and microsatellites did not differ significantly from those obtained only with SNPs. The estimated average African ancestry in X chromosomes from Cape Verde was 0.65, in accordance with historical data indicating that mixed unions were highly asymmetric and almost invariably involved European males and African females. Average individual African ancestry estimates in each island were 0.81 in Santiago, 0.68 in Santo Antão, 0.64 in São Vicente, 0.62 in São Nicolau, 0.61 in Fogo and 0.59 in Boavista. The differences between the islands globally reflect the properties of the several peopling cycles of Cape Verde, although admixture levels in Fogo are somewhat higher than expected given the similarities between Fogo and Santiago's plantation systems. The variation in admixture proportions across islands is in qualitative accordance with the admixture levels calculated with the Y chromosome and several autosomal markers. However, the X-chromosome African-ancestry estimates are lower than expected in simple panmixy models. The proportion of ancestry switches in the X chromosome was found to be highly correlated ($R^2=0.86$) with the recombination fractions, and a clear drop in recombination was found close to the centromere, where a contiguous ~14 Mb fragment is likely to preserve the original ancestry of each chromosome. We used this finding to explore a method for estimating the time since admixture and obtained estimates ranging from ~ 330 to ~500 years, compatible with the historical data. The finding of significant correlations between the size of European and African chromosome fragments and the observed levels of admixture, together with the evolutionary dynamics of admixture in African and European chromosomes, indicate that the admixture process has still not reached equilibrium. Our results represent the first attempt to characterize admixture in Cape Verde using an analytical approach with higher complexity.

1. Introdução

A formação de populações miscigenadas a partir de populações ancestrais geneticamente diferenciadas é um processo demográfico que pode ser utilizado para modelar muitas questões de microevolução humana [1]. Por exemplo, a influência de populações de origem neolítica ou paleolítica no património genético dos europeus, ou as consequências do contacto entre populações de língua Bantu e não-Bantu durante a expansão da agricultura em África, são problemas que podem ser estudados a partir do conceito de miscigenação [2, 3]. Na maior parte dos casos, porém, a miscigenação é tão antiga que não é possível identificar as populações parentais, nem calcular o tempo que decorreu desde que essas populações começaram a fundir-se. As populações que se formaram na sequência de acontecimentos históricos recentes, pelo contrário, ainda conservam os sinais dos fenómenos de miscigenação que as originaram e são experiências naturais que podem ser usadas para estudar não só os processos de miscigenação, mas também a biologia de características complexas com diferentes prevalências nos grupos parentais [4, 5].

O arquipélago de Cabo Verde, devido à sua localização numa região estratégica para o tráfico de escravos capturados na África Ocidental, reúne condições quase ideais para analisar processos de miscigenação entre populações africanas e europeias. O objectivo desta tese é contribuir para descrever e interpretar a miscigenação humana no arquipélago de Cabo Verde através do estudo da distribuição de segmentos de ancestralidade europeia e africana no cromossoma X. O trabalho faz parte de um projecto de colaboração com a Universidade de Cabo Verde (UniCV), que procura caracterizar a estrutura genética da população do arquipélago para ajudar a reconstruir a história do seu povoamento e elucidar a base genética de características complexas com interesse antropológico (como a pigmentação cutânea) ou biomédico (hipertensão e obesidade). Para além da análise do cromossoma X, de que agora se apresentam resultados preliminares, a caracterização genética de Cabo Verde tem envolvido um número muito diversificado de marcadores moleculares localizados noutras regiões genómicas, incluindo os autossomas, e os cromossomas Y e mitocondrial. Alguns dos resultados obtidos nesses estudos serão aqui apresentados para efeitos comparativos.

1. 1. História do povoamento de Cabo Verde

O arquipélago de Cabo Verde, situado no oceano Atlântico, entre o trópico de câncer e o equador, a cerca de 450 km da costa senegalesa, divide-se tradicionalmente em dois grupos: Barlavento – constituído pelas ilhas de Santo Antão, S. Vicente, Santa Luzia, São Nicolau, Sal e Boa Vista – e Sotavento, integrado pelas ilhas de Maio, Santiago, Fogo e Brava (Figura 1). Para Correia e Silva [6], esta divisão baseada em critérios de náutica veleira não reflecte bem as características ecológicas que influenciaram o povoamento e as actividades económicas nas várias ilhas. São Vicente e Santa Luzia, no Norte, e Sal, Boavista e Maio, no Oriente, são ilhas planas e secas, nas quais nunca se desenvolveu a agricultura. Santiago, a ilha maior, tem bons portos naturais e nascentes de água doce. A ilha do Fogo, apesar da reduzida disponibilidade de água, é alta e húmida e tem condições para agricultura de sequeiro. As ilhas da Brava, no Sul, e Santo Antão e São Nicolau, no Norte, são montanhosas e ricas em recursos hídricos, com boas condições para agricultura de sementeira e regadio [6-9]. Com estas características, há ilhas do Barlavento que podem ser agrupadas com ilhas do Sotavento (por exemplo Brava e Santo Antão), e ilhas do mesmo espaço geográfico que devem ser distinguidas (por exemplo Santiago e Maio).

Quando foi descoberto nos inícios da década de 1460, o arquipélago de Cabo Verde era despovoado. De acordo com Correia e Silva [7], é possível distinguir três ciclos de ocupação do território. O primeiro ciclo compreende os séculos XV e XVI e corresponde à ocupação de Santiago em 1462, seguida da do Fogo, entre 1480 e 1493. Nesta altura a população era constituída por colonos de origem europeia (na maioria portugueses) e escravos provenientes da região situada entre a Senegâmbia e a Serra Leoa, que eram trocados por algodão, cavalos e outros produtos levados de Santiago e do Fogo [6, 10]. Os escravos africanos, que excediam largamente em número os colonos europeus, tinham dois destinos principais]. A maioria era exportada para as Antilhas, América Central e Brasil. Os restantes permaneciam em Cabo Verde e eram divididos em “escravos rurais” e “escravos domésticos”. É provável que a mestiçagem em Cabo Verde tenha resultado de uniões entre escravas domésticas e colonos europeus, já que a presença de mulheres europeias era rara, pelo menos nos estados iniciais da colonização [10]. A língua de Cabo Verde é um dos legados culturais mais significativos deste processo de miscigenação e, de acordo com as fontes históricas disponíveis, o seu uso já estaria generalizado por volta de 1540 [10].

Durante o primeiro ciclo de povoamento as restantes ilhas do arquipélago eram usadas para criação intensiva de gado- que podia ser deixado livre durante grande parte do ano – mantendo-se praticamente desabitadas nos séculos XV e XVI [8]. As ilhas do Sal, Boavista e Maio, localizadas na região oriental do arquipélago (Figura 1), tinham depósitos naturais de sal que não necessitava de grande força de trabalho para ser colhido. A partir do século XVII, a intrusão de várias potências europeias na costa ocidental de África levou ao declínio do monopólio escravagista e desencadeou a ocupação das ilhas da Brava, no Sul, e Santo Antão e São Nicolau, no Norte (Figura 1). É provável que, com excepção de Santo Antão, a ocupação tivesse sido feita integralmente por camponeses livres, na sua maioria mestiços, atraídos pelas boas condições para agricultura daquelas ilhas [7]. Neste processo, a quase total ausência de escravos na força de trabalho, a pequena dimensão das parcelas de terra cultivada, e a diversificação de culturas agrícolas, contrastavam fortemente com o sistema de plantação que prevaleceu no Fogo e em Santiago nos séculos XV e XVI e criaram uma distinção nítida entre Cabo Verde e outras sociedades crioulas nascidas do tráfico de escravos [7, 11].

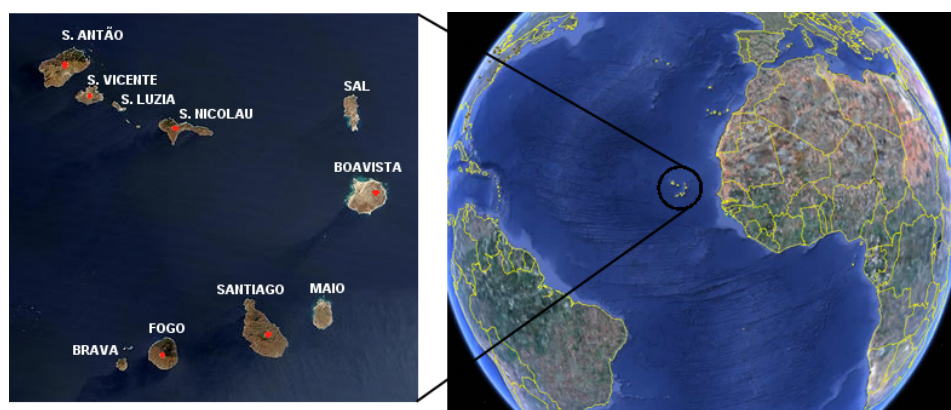


Fig. 1: Localização do arquipélago de Cabo Verde e das suas ilhas.

O terceiro ciclo de povoamento corresponde à tentativa de ocupação continuada das ilhas de São Vicente e Santa Luzia durante o século XVIII (Figura 1). Contudo, devido à aridez destas ilhas, todas as tentativas foram mal sucedidas. Só no século XIX, com o início da navegação comercial e a abertura do porto da cidade do Mindelo, é que São Vicente foi efectivamente ocupada, tornando-se rapidamente na segunda ilha mais povoada do arquipélago, a seguir a Santiago [7].

As ilhas orientais do Sal, Boavista e Maio, mantiveram-se à margem dos três ciclos de povoamento (Figura 1). Escassamente povoadas durante o primeiro ciclo e sem as boas condições agrícolas das ilhas ocupadas no segundo ciclo, as três ilhas continuaram a ser usadas para a recolha de sal e criação de gado até à abertura no século XVIII dos portos da Boavista e Maio aos navios ingleses e norte-americanos [7]. Mesmo assim, estas ilhas nunca se tornaram centros demográficos importantes e em conjunto representam apenas cerca de 5% da população total do arquipélago.

1. 2. Miscigenação e ancestralidade de fragmentos cromossómicos

No estudo de populações miscigenadas, um dos principais parâmetros a calcular é a contribuição genética média das populações ancestrais (ou parentais) para a nova população híbrida [12]. O princípio deste cálculo é relativamente simples porque as frequências alélicas nas populações miscigenadas são dadas pela média das frequências alélicas das populações parentais, ponderada pelas contribuições relativas dessas populações para a nova população híbrida. Nestas condições, será possível calcular as proporções ancestrais simplesmente com base nas frequências alélicas das populações parentais e das populações híbridas [12].

Os estudos pioneiros sobre miscigenação em populações humanas, quase todos centrados em populações afro-americanas, basearam-se neste princípio. Por exemplo, Reed [13], usando marcadores genéticos clássicos, observou uma grande variação (4-26%) nos níveis de ancestralidade europeia em populações afro-americanas de diferentes áreas dos Estados Unidos. Num estudo mais recente, Parra et al. [14] confirmaram esta tendência, usando uma bateria de marcadores moleculares com grandes diferenças de frequências alélicas ($>0,6$) entre populações europeias e africanas. Ao maximizar a divergência entre populações ancestrais, estes *loci*, comumente designados por marcadores informativos para a ancestralidade (AIMs-*Ancestral Informative Markers*), aumentam muito a precisão das estimativas de miscigenação, mesmo quando se usa um número relativamente pequeno de marcadores.

As proporções de ancestralidade numa população são estimativas médias que nos dão informações importantes sobre os níveis de miscigenação, mas nada nos dizem sobre o modo como a miscigenação está distribuída nos indivíduos da população híbrida [15]. Por exemplo, numa população híbrida constituída por igual número de descendentes de dois grupos A e B que não se cruzam entre si é possível calcular uma contribuição média de A e B de 50%. No entanto, se se calculassem os níveis individuais de miscigenação nessa população, verificar-se-ia que metade dos indivíduos teria a totalidade da sua ancestralidade na população A e metade na população B, o que corresponderia à ausência total de panmixia na população híbrida. Se, pelo contrário, cada indivíduo apresentasse proporções variáveis de ancestralidade em A e B, e a média populacional das contribuições de A e B fosse de 50%, já seria mais provável que os cruzamentos se fizessem ao acaso na população híbrida, irrespectivamente da proveniência dos ancestrais de cada indivíduo. Por outras palavras, é possível obter níveis médios de miscigenação idênticos em populações com

dinâmicas de miscigenação muito diferentes, que só podem ser distinguidas quando se conhece a heterogeneidade dos níveis de miscigenação individuais.

Os níveis de miscigenação individuais podem ser ainda usados para analisar correlações entre ancestralidade e susceptibilidade a doenças com diferenças de prevalência significativas nas populações parentais. Por exemplo, a diabetes de tipo 2 tem uma prevalência significativamente mais elevada em algumas populações ameríndias do que em populações de origem europeia, sem que se saiba ao certo se essa diferença resulta de factores genéticos ou ambientais. Para responder a esta questão, Williams et al [16] avaliaram a correlação entre os níveis de ancestralidade europeia e factores de risco à diabetes numa população miscigenada de índios Pima, tendo concluído que os indivíduos não-diabéticos tinham o dobro da ancestralidade europeia dos diabéticos. Um resultado deste tipo indica que há uma forte probabilidade de as diferenças de prevalência da diabetes serem causadas por factores genéticos de susceptibilidade dos índios Pima. No entanto, esta conclusão só é válida se se puder garantir que não há factores ambientais associados às diferenças nos níveis individuais de miscigenação.

O cálculo dos níveis individuais de miscigenação tem uma variância muito grande e exige mais marcadores do que a estimativa das proporções de ancestralidade média numa população. No entanto, a possibilidade de se usarem baterias de AIMS, a evolução das técnicas de genotipagem e a disponibilidade crescente de métodos computacionais, tem vindo a facilitar a análise da distribuição das proporções de ancestralidade individuais e não apenas dos seus valores médios [15, 17, 18].

Há ainda um outro nível de complexidade nos estudos de miscigenação que corresponde à determinação da ancestralidade de diferentes porções do genoma de cada indivíduo. Uma das consequências mais interessantes da miscigenação é a formação de cromossomas híbridos nos quais há segmentos alternados com origem em cada uma das populações parentais (Figura 2). Como os cromossomas híbridos resultam do processo de recombinação entre cromossomas das populações parentais, o estudo do tamanho e do número de segmentos cromossómicos com diferentes origens permite estimar a intensidade da miscigenação, a idade e a dinâmica dos fenómenos de miscigenação, bem como a distribuição da recombinação ao longo de um cromossoma [17-22]. Por outro lado, ao contrário da variação genética em regiões não recombinantes do genoma, como o DNA mitocondrial (mtDNA) ou o cromossoma Y, tradicionalmente usada para caracterizar a história das populações, os cromossomas híbridos são gerados por recombinação na população miscigenada e dão-nos informação sobre a dinâmica do processo de miscigenação.

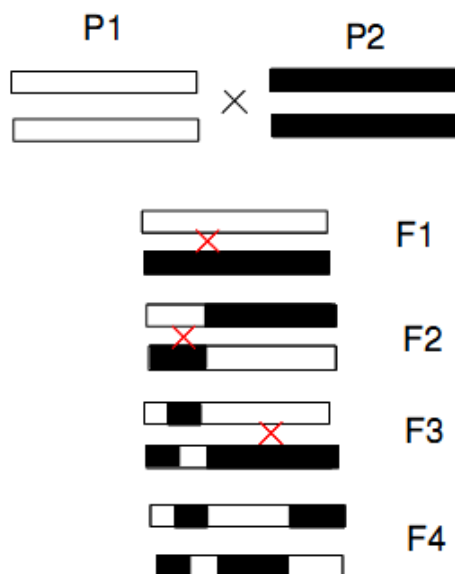


Figura 2: Esquema do processo e formação de segmentos cromossômicos alternados com origem em duas populações parentais diferentes (P1 e P2) ao longo de sucessivas gerações (F1, F2....F4) numa população miscigenada. Os pontos de recombinação estão indicados a vermelho (Modificado de Ungerer et al. [19]).

Outra aplicação importante do estudo de segmentos cromossômicos com distintas proveniências parentais é o chamado mapeamento por miscigenação (AM, *admixture mapping*). Nesta abordagem, procura-se identificar nas populações miscigenadas os genes responsáveis por características complexas com diferentes prevalências nas populações parentais. No caso da diabetes tipo 2 acima mencionado, por exemplo, o passo seguinte ao estudo de correlação entre ancestralidade individual e risco diabetes seria fazer uma análise de mapeamento por miscigenação com base na distribuição da proveniência parental de segmentos cromossômicos. O princípio em que se baseiam os estudos de AM é o de que nas regiões cromossômicas com variantes que contribuem para dada característica estará sobrerrepresentada a população parental onde essa característica é mais frequente [5, 23] .

Actualmente existem vários métodos para inferir a ancestralidade de diferentes regiões cromossômicas em populações híbridas [5, 17, 20, 23]. Quando os marcadores genéticos são seleccionados aleatoriamente, este tipo de inferência requer uma grande quantidade de *loci* densamente distribuídos ao longo de cada cromossoma, o que encarece substancialmente os estudos e limita a sua aplicabilidade. No entanto, se se usarem AIMs, a observação de um dado alelo num cromossoma híbrido deixará poucas dúvidas sobre a população de que provém e a delimitação de segmentos cromossômicos poderá ser obtida com um número razoavelmente menor de *loci* [5, 17, 20, 23]. Ainda assim, os métodos de inferência de blocos cromossômicos dependem criticamente da capacidade de realizar uma reconstrução correcta dos arranjos haplotípicos a partir dos genótipos observados nas populações parentais [5, 17, 20, 23].

Em resumo pode dizer-se que há três níveis diferentes de complexidade nos estudos de miscigenação: o nível mais simples corresponde à estimativa das contribuições médias das populações parentais para a população miscigenada; o nível seguinte consiste na avaliação das proporções de ancestralidade individuais no seio das populações miscigenadas; o terceiro nível consiste na identificação de fragmentos cromossômicos com diferentes

ancestralidades no genoma de cada indivíduo. Partindo do terceiro nível é possível calcular os níveis individuais de miscigenação e a partir desses níveis estimar os níveis médios de miscigenação de uma população. No entanto, não é possível percorrer o caminho inverso e obter miscigenações individuais a partir de níveis médios populacionais, nem delimitar blocos cromossômicos com origem nas diferentes populações parentais a partir do cálculo da miscigenação individual.

O estudo que agora se apresenta é uma tentativa de analisar pela primeira vez os níveis de miscigenação da população caboverdiana com o nível analítico de maior complexidade. Para contornar as principais dificuldades da reconstrução de blocos cromossômicos, fez-se uma escolha criteriosa de marcadores genéticos com grandes diferenças de frequências alélicas entre populações africanas e europeias (as populações parentais de Cabo Verde) e utilizou-se, como modelo, o cromossoma X. O facto de o cromossoma X ser diplóide nas mulheres e haplóide nos homens permite fazer a inferência haplotípica nos indivíduos do sexo masculino sem recurso a métodos estatísticos e sem incerteza. A utilização do cromossoma X tem ainda duas vantagens adicionais: a inexistência de recombinação nos homens desacelera o processo de fragmentação dos segmentos cromossômicos parentais e facilita a definição de segmentos com uma densidade de marcadores relativamente baixa; por outro lado, a alternância do cromossoma X ao longo das gerações entre o sexo masculino e o sexo feminino permite analisar as assimetrias no comportamento demográfico de homens e mulheres durante o processo de miscigenação [24].

2. Material e Métodos

2. 1. Composição da amostra

A amostra é composta por 210 indivíduos não aparentados do sexo masculino provenientes das seis ilhas mais povoadas do arquipélago de Cabo Verde: Santiago (n=25), Fogo (n=64), Boavista (n=21), Santo Antão (n=42), São Nicolau (n=34), São Vicente (n=24). Escolheram-se apenas indivíduos do sexo masculino para facilitar a inferência dos arranjos haplotípicos que permitem identificar segmentos cromossômicos de diferente proveniência.

2. 2. Marcadores do cromossoma X

A definição de segmentos do cromossoma X com ancestralidade europeia e africana foi feita com 37 polimorfismos nucleotídicos simples (SNPs, *Single Nucleotide Polymorphisms*) e 14 microssatélites situados a intervalos regulares ao longo de todo o cromossoma (1,8 Morgan). Os SNPs têm diferenças de frequências alélicas entre populações africanas e europeias maiores que 60% (i.e são AIMs) e um espaçamento médio de 4,86 cM (centimorgan), com um mínimo de 0,01 e um máximo de 22,96 (Tabela 1). Os microssatélites estão localizados em regiões flanqueadas pelos SNPs, com um espaçamento médio de 12,86 cM (mínimo de 0,82; máximo 38,29), e não foram seleccionados em função das diferenças de frequências alélicas nas populações parentais (Tabela 2). No conjunto há 1 marcador (SNP ou microssatélite) por cada 3,53 cM (Figura 3).

2. 3. Delimitação de fragmentos cromossômicos com proveniência africana ou europeia

A identificação da proveniência dos segmentos cromossômicos foi feita com o algoritmo Bayesiano disponível no programa STRUCTURE [17], usando o modelo de ligação factorial (*linkage model*) com fase haplotípica conhecida, assumindo duas populações parentais (K=2) com frequências alélicas desconhecidas. As distâncias genéticas entre marcadores foram calculadas a partir do mapa disponibilizado pelo projecto HapMap (HapMap rel21/NCBI rel 35; ver Tabelas 1 e 2). Foram feitas 25 corridas por cada indivíduo com 600 000 réplicas por corrida após um período inicial (*burn-in*) de 200 000 réplicas. Nestas condições obtiveram-se sempre resultados quase idênticos entre as diferentes corridas. No trabalho apresentam-se as médias dos valores obtidos nas 25 corridas.

Tabela1: Características dos SNPs analisados

SNP ¹	Frequência alélicas ²		δ^3	Posição no mapa físico (bp) ⁴	Posição no mapa genético (cM) ⁵
	África	Europa			
rs211644	0,92	0,285	0,635	2872721	0,846193229
rs723259	0,72	0,041	0,679	9799198	14,83482967
rs1474458	0,736	0,89	0,647	10957061	16,09717636
rs3747295	0,067	0,943	0,876	17656165	25,91105465
rs6633170	0,949	0,221	0,728	18847131	27,49713619
rs5955757	0,965	0,02	0,945	19283760	28,90164201
rs5973335	0,972	0,367	0,605	29854423	42,22743057
rs1011759	0,081	0,731	0,65	35242142	54,95888792
rs5963039	0,798	0,103	0,695	38234863	58,7220049
rs7887756	0,705	0,077	0,628	44668666	70,04445766
rs5952682	0,312	0,923	0,611	44851739	70,22354659
rs6609827	0,659	0	0,659	48818318	74,74679958
rs7061449	0,129	0,994	0,865	54036020	78,16264539
rs7886355	0,944	0,038	0,906	62875973	78,82658276
rs35972027	0,112	0,923	0,811	65083915	79,08961553
rs699863	0,016	0,814	0,798	65369520	79,13808308
rs492933	0,99	0,373	0,617	67180569	79,42729874
rs2209549	0,86	0,141	0,719	73875354	85,48873103
rs4892539	0,767	0,042	0,725	74215399	85,50371295
rs234256	0,751	0,012	0,739	75870206	85,66310316
rs1015736	0,204	0,957	0,753	85575340	89,24161814
rs764581	0,975	0,308	0,667	95327205	100,1830523
rs1020903	0	0,903	0,903	99040050	104,1408202
rs768312	0,694	0,006	0,688	109558040	110,2750734
rs992864	0,067	0,943	0,876	110387239	110,6170019
rs5988008	0,152	0,818	0,666	114864565	116,3979857
rs714749	0,272	0,938	0,666	117164560	120,3415618
rs781045	0,71	0,076	0,634	119016317	123,3143106
rs766751	1	0,378	0,622	126496367	136,4160082
rs909657	0,857	0,088	0,769	128611248	138,3104305
rs1908816	0,978	0,16	0,818	132808721	140,6889579
rs12687833	0,739	0,064	0,675	138669673	147,1258707
rs764536	0,966	0,078	0,888	147767215	170,0898577
rs758439	0,966	0,287	0,679	147872587	170,3604779
rs732604	0,275	0,938	0,663	148190759	170,8084588
rs5201	0,95	0,266	0,684	152825187	180,4014954
rs12557310	0,955	0,269	0,686	154388892	180,9830504

1. SNPs seleccionados a partir da base de dados disponibilizada por Smith et al. [25]; 2. Frequências de um dos alelos dos SNPs disponibilizadas por Smith et al. [25] , calculadas com base em 198 indivíduos da África Ocidental (Nigéria (36), Gana (50), Camarões (28) e Senegal (84)) e 119 indivíduos americanos de origem europeia. 3. Valor absoluto da diferença entre frequências alélicas na Europa e em África; 4. De acordo com a base de dados disponibilizada pela University of Califórnia at Santa Cruz (UCSC), versão de Março de 2006b(NCBI36/Hg18); 5. De acordo com informação disponibilizada pelo projecto HapMap (HapMap rel21/NCBI rel 35).

Tabela 2: Características dos STRs analisados

STR ¹	Sequência repetida	Posição no mapa físico ² (bp)		Posição no mapa genético ³ (cM)	
		Início	Fim	Início	Fim
GATA175D03	(TAGA) _n	15233383	15233816	23,76185503	23,76192683
ATA28C05	(AAT) _n	23154419	23154739	36,16277897	36,16789254
GATA186D06	(CTAT) _n	32010306	32010626	48,62213039	48,62225123
GATA144D04	(TCTA) _n	44898235	44898750	70,26400851	70,26403742
GATA72E05	(GATA) _n	64571993	64572357	78,94848688	78,94853078
AGAT104	(AGAT) _n	72866354	72866476	85,27589261	85,27589268
GATA31D10	(TATC) _n	78566980	78567290	86,10516025	86,10516044
GATA31F01	(AGAT) _n	95335896	95336212	100,1831579	100,1831616
GATA10C11	(TCTA) _n	107367696	107368054	108,4443496	108,445204
GATA172D05	(AGAT) _n	113061267	113061309	113,8457319	113,8457381
GATA48H04	(ATAG) _n	116391039	116391622	118,7268187	118,727055
GATA198A10	(AGAT) _n	122001710	122001775	127,4006749	127,400751
AAAT112	(TTAT) _n	134655314	134655363	142,4588955	142,4588972
TTTA062	(TTTA) _n	153184723	153184773	180,7502742	180,750277

1. STRs seleccionados a partir da base de dados disponibilizada pelo Marshfield Center for Medical Genetics.
2. De acordo com a base de dados disponibilizada pela UCSC, versão de Março de 2006 (NCBI36/Hg18)
3. De acordo com informação disponibilizada pelo projecto HapMap (HapMap rel21/NCBI rel 35)

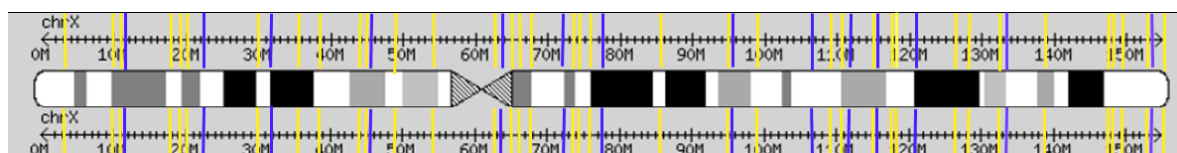


Figura 3: Ideograma do cromossoma X com localização física aproximada (em megabases, M) dos marcadores usados para inferir segmentos cromossômicos de proveniência europeia e africana. A localização dos SNPs está indicada com barras verticais amarelas e a dos microssatélites com barras verticais azuis.

2. 4. Genotipagem de SNPs e microssatélites

A genotipagem dos SNPs foi realizada no Instituto Gulbenkian de Ciência com métodos multiplex automatizados que permitem a análise simultânea de vários *loci* recorrendo a tecnologia iPLEX™.

A genotipagem dos microssatélites foi feita por amplificação em reacção em cadeia da polimerase (PCR, *Polymerase Chain Reaction*) em três reacções multiplex com separação dos fragmentos num sequenciador automático ABI 3130. A análise dos fragmentos foi feita com o programa *GeneMapper Analysis Software V. 4.0*. Na figura 4 mostra-se um exemplo da resolução alcançada numa das reacções multiplex para seis microssatélites.

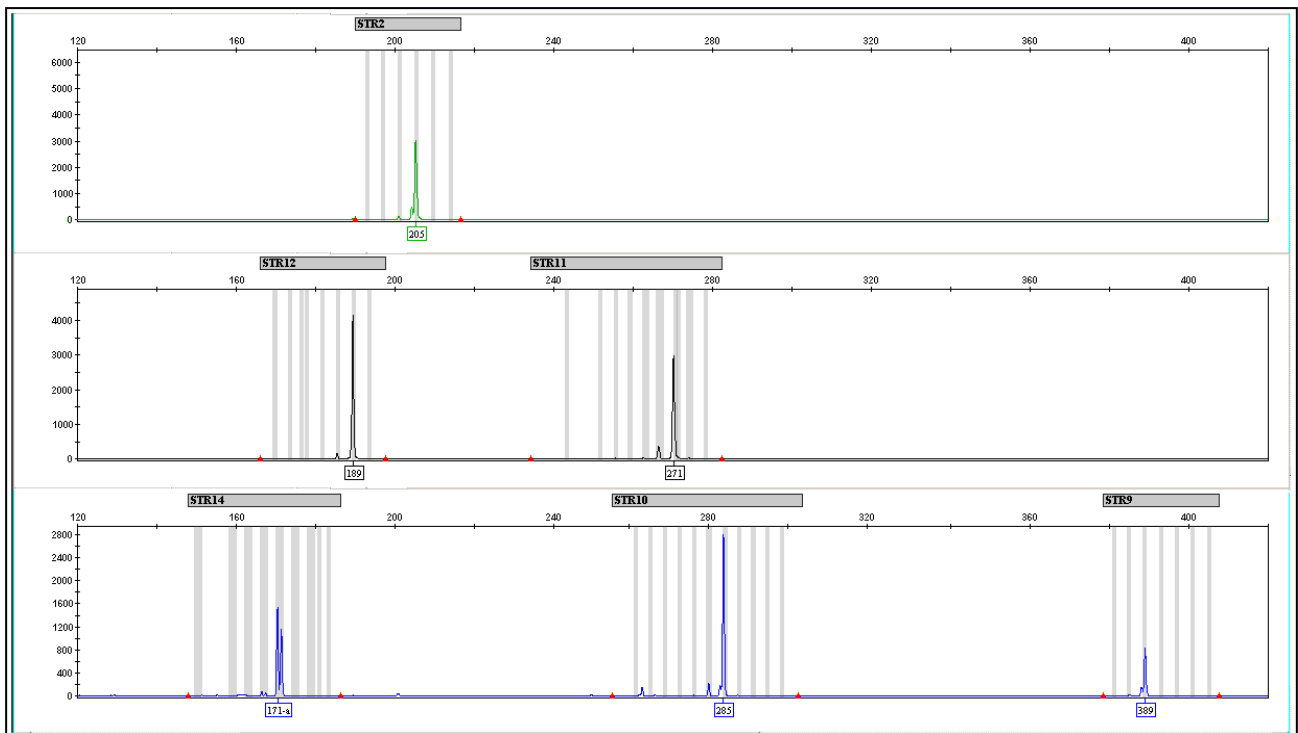


Figura 4: Exemplo da análise, no GeneMapper (Applied Biosystems), de um dos multiplex estudados incluindo seis STRs. A detecção foi feita com o sequenciador automático ABI3130. A marcação de fluorescência usada em cada marcador é indicada pelas cores azul (6-FAM), verde (VIC) e NED (preto).

3. Resultados e Discussão

3.1. Resultados obtidos com SNPs

O principal objectivo do modelo de ligação (*linkage model*) disponibilizado no programa STRUCTURE é determinar a proveniência dos alelos de cada marcador genético na população miscigenada, tendo em conta a informação dos *loci* vizinhos. Como tanto a capacidade de imputar a proveniência de um alelo isolado como a correlação entre alelos de *loci* diferentes dependem das diferenças de frequências alélicas das populações parentais, os AIMs são os marcadores de eleição para a identificação de segmentos cromossômicos com diferente ancestralidade. Por estas razões, optou-se por apresentar em primeiro lugar os resultados obtidos apenas com SNPs que, como já foi dito, têm grandes diferenças de frequência entre populações africanas e europeias e se comportam como AIMs (Tabela 1). Numa segunda fase, apresentar-se-ão os resultados da análise simultânea de SNPs e microssatélites.

3.1.1 Distribuição de segmentos cromossômicos com ancestralidade africana e europeia

Na figura 5 apresentam-se, para efeitos ilustrativos, as probabilidades de origem africana dos alelos dos 37 SNPs do cromossoma X, calculadas com o programa STRUCTURE para os indivíduos com maior ancestralidade africana (Fig. 5A) e europeia (Figura 5B) da amostra (**ver Anexo 1**). No indivíduo com ancestralidade africana, natural da ilha de Santiago, os alelos da grande maioria dos loci (36/37=97%) têm uma probabilidade de origem africana superior a 75%. Há apenas um marcador cujo alelo não pode ser imputado a África com essa probabilidade, embora por uma diferença tangencial (Figura 5A). Com estes resultados, é razoável concluir que este cromossoma X forma um só bloco de origem africana que não incorporou fragmentos de origem europeia por recombinação (Figura 5C). No indivíduo em que predomina a ancestralidade europeia, natural de Santo Antão, 81% (30/37) dos alelos têm probabilidades de origem africana inferiores a 25%, o que corresponde, logicamente, a probabilidades de origem europeia superiores a 75%. Em três loci os alelos têm probabilidades de origem entre 25% e 75% e há quatro marcadores consecutivos cujos alelos têm uma probabilidade de origem africana superior a 75% (Figura 5A). Este resultado sugere que houve a integração de pelo menos um bloco de origem africana num cromossoma de origem europeia (Figura 5 D).

A percentagem de situações em que não foi possível imputar a ancestralidade de um alelo com uma probabilidade superior a 75% - correspondente a probabilidades de ancestralidade ente 25% e 75% - é uma medida da incerteza associada ao método de imputação e às condições em que foi utilizado (incluindo o tipo de marcadores e a sua densidade). Nas condições aqui usadas, das 7770 imputações (37 marcadores x 210 indivíduos) 975 (12,5%) tiveram probabilidades de ancestralidade africana ente 25% e 75%, correspondentes a situações de imputação incerta. Este valor é muito mais baixo do que a percentagem de 86% de incerteza obtida quando se usaram apenas 14 microssatélites (resultados não mostrados), justificando a opção de apenas se apresentarem os resultados relativos a SNPs ou à análise simultânea de SNPs e microssatélites.

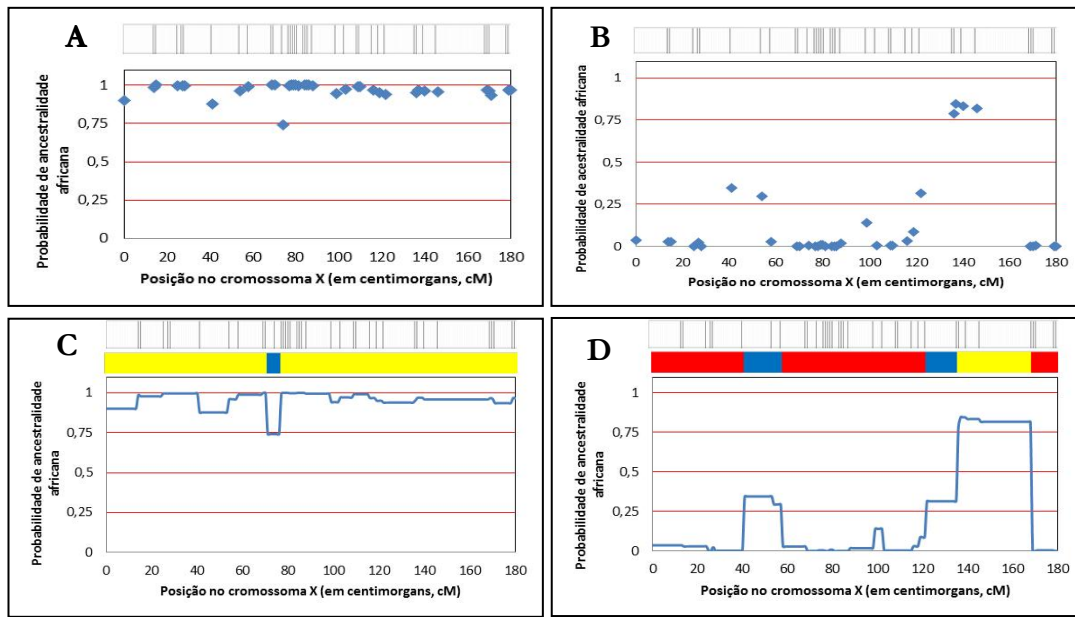


Figura 5: **A e B**, Probabilidades de origem africana dos alelos de 37 SNPs localizados no cromossoma X dos indivíduos com maior ancestralidade africana (**A**) e europeia (**B**) da amostra. **C e D**, Representação de blocos de ancestralidade africana e europeia no cromossoma X dos indivíduos representados em **A** e **B**, respectivamente. Os blocos com probabilidade de origem africana >75% estão representados a amarelo e os blocos com probabilidade de origem europeia >75% a vermelho. Os blocos com valores de probabilidade entre 25% e 75% estão representados a azul. Em cada painel, o gráfico de barras representa a posição relativa de cada SNP no cromossoma X.

A incerteza das imputações não parece estar distribuída ao acaso (Figura 6). Há marcadores, como o SNP rs597335 (na posição relativa nº7), em que a percentagem de imputações incertas é claramente superior à média (50% vs 12,5%) e outros - entre os quais se destacam os loci pericentroméricos rs7886355, rs35972027, rs699863 e rs492933 (loci nº 14, 15, 16 e 17, ver figura 3) – em que a incerteza na imputação é muito baixa (Figura 6A). Um dos factores que pode explicar esta heterogeneidade é a variação na amplitude das diferenças entre frequências alélicas nas populações parentais em cada marcador (δ , ver Tabela 1). De facto, como se mostra na Figura 6B, há uma correlação negativa ($R^2=0,127$) entre os valores de δ e a percentagem de indefinição nas imputações de ancestralidade. Outro factor que pode influenciar a precisão das imputações é a distância genética em cM entre cada marcador e os marcadores adjacentes, tendo em conta que a correlação entre loci consecutivos, ou desequilíbrio de ligação (LD, *linkage disequilibrium*), é usada pelo algoritmo do programa STRUCTURE para identificar a ancestralidade. A figura 6 C ilustra precisamente este efeito e mostra que a percentagem de indefinição está positivamente correlacionada ($R^2=0,183$) com soma das distâncias de cada marcador aos marcadores adjacentes.

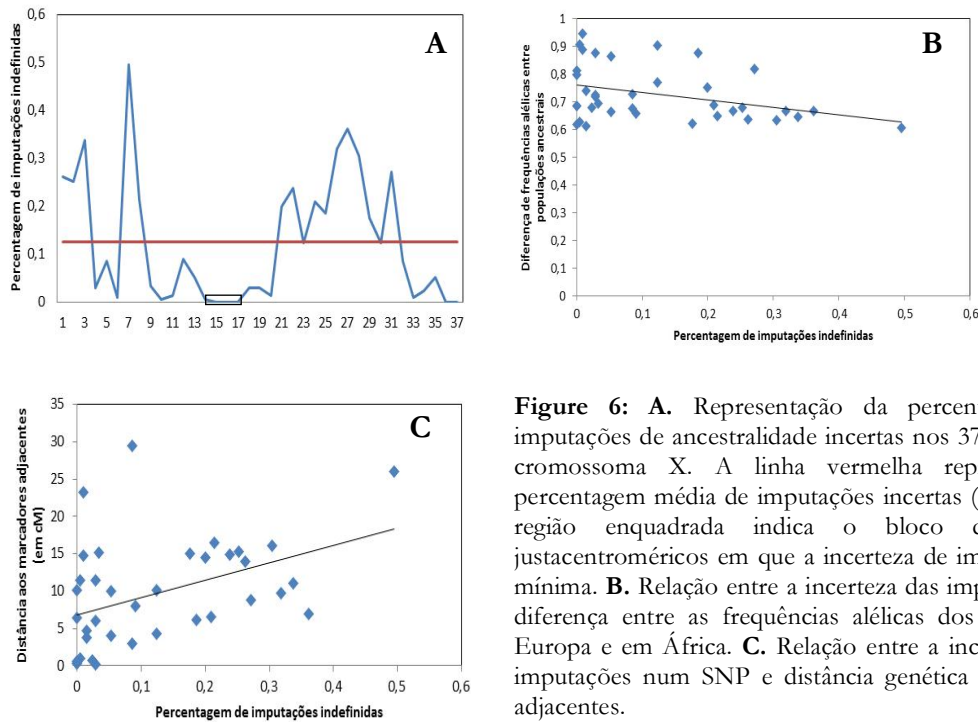


Figure 6: **A.** Representação da percentagem de imputações de ancestralidade incertas nos 37 SNPs do cromossoma X. A linha vermelha representa a percentagem média de imputações incertas (12,5%). A região enquadrada indica o bloco de SNPs justacentroméricos em que a incerteza de imputação é mínima. **B.** Relação entre a incerteza das imputações e diferença entre as frequências alélicas dos SNPs na Europa e em África. **C.** Relação entre a incerteza das imputações num SNP e distância genética aos SNPs adjacentes.

Outra fonte de incerteza é a baixa densidade de marcadores ao longo do cromossoma. O processo mais simples de representar blocos com diferente ancestralidade a partir de um número finito de marcadores é assumir que cada bloco ocupa toda a região compreendida entre loci consecutivos, tal como representado nas figuras 5 C e D. Se, por exemplo, num cromossoma com dois marcadores consecutivos ($M1$ e $M2$) o alelo do marcador $M1$ tiver proveniência europeia e o alelo do marcador $M2$ tiver proveniência africana, considera-se que há um bloco de ancestralidade europeia entre $M1$ e $M2$ e que há um bloco com ancestralidade africana que começa em $M2$. No entanto, é inteiramente possível que a transição de ancestralidade entre os dois blocos consecutivos - a que se chamará aqui *sutura*- tenha ocorrido em qualquer posição situada entre $M1$ e $M2$ [19]. Nestas condições, a localização exacta das suturas e a estimativa do tamanho dos blocos parentais será tanto mais difícil quanto menor for a densidade de marcadores utilizados para os inferir. Por outro lado, os blocos com tamanho inferior ao da distância entre marcadores nunca poderão ser detectados [17, 19]. Estas limitações estão contudo minoradas no cromossoma X dado que, como se referiu na introdução, a ausência de recombinação nos homens permite manter a integridade dos blocos parentais durante maior número de gerações.

Na figura 7 exemplifica-se a estrutura em mosaico de cromossomas X das ilhas de Santiago e São Vicente que resultou da recombinação entre cromossomas de origem africana ou europeia. Devido às limitações acima discutidas, e ao contrário das figuras 5 C e D, não se procurou apresentar a extensão dos segmentos com distinta ancestralidade mas apenas a proveniência parental inferida para cada marcador com base nas probabilidades calculadas pelo programa STRUCTURE. Mesmo assim, este tipo de representação ilustra bem a alternância das suturas ao longo de cada cromossoma e mostra que os cromossomas X da ilha de São Vicente tem maior número de segmentos de origem europeia do que os de Santiago.

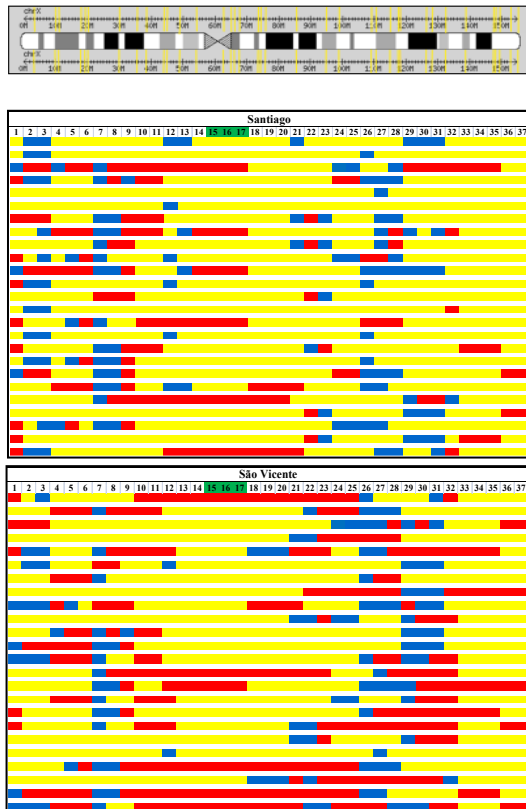


Figura 7: Representação esquemática da ancestralidade africana (amarelo) ou europeia (vermelho) dos alelos de 37 SNPs do cromossoma X. A cor azul representa imputações incertas. Cada cromossoma X é representado por uma linha. No painel de cima mostra-se a localização física dos SNPs no cromossoma X. O painel seguinte mostra cromossomas da ilha de Santiago. O painel de baixo mostra cromossomas da ilha de São Vicente. Os SNPs sublinhados a verde são os *loci* justacentroméricos em que não há incerteza na imputação de ancestralidade.

3.1.2 Estimativas de miscigenação

3.1.2.1 Distribuição da miscigenação individual no arquipélago

A percentagem de alelos com ancestralidade europeia ou africana nos 37 SNPs de cada cromossoma X é uma medida da contribuição relativa das populações parentais para esse cromossoma. Esta medida equivale a uma estimativa individual de ancestralidade baseada no cromossoma X e a sua média dá-nos a contribuição das populações parentais para os cromossomas X da população do arquipélago. Para imputar a ancestralidade dos alelos observados em cada um dos 37 SNPs, usaram-se dois critérios. No primeiro critério, descartaram-se as imputações com maior incerteza ($>25\%$ e $<75\%$) e considerou-se que um alelo tinha origem europeia ou africana quando a sua probabilidade de origem ancestral era superior a 75% . No segundo critério, relaxou-se o nível de confiança das imputações e considerou-se que um alelo provinha da Europa ou África sempre que a sua probabilidade de origem ancestral fosse superior a 50% . Os resultados obtidos com os dois critérios estão fortemente correlacionados ($R^2=0,99$). Em ambos os casos a estimativa média da contribuição africana para os cromossomas X de Cabo Verde foi de 65% (Figura 8A e B) e não se detectaram diferenças significativas entre as duas distribuições com um teste de Kolmogorov-Smirnov para duas amostras ($P=0,728$). Assim, a menos que seja explicitamente referido o contrário, daqui para a frente apenas se apresentarão resultados referentes ao critério de imputação mais relaxado.

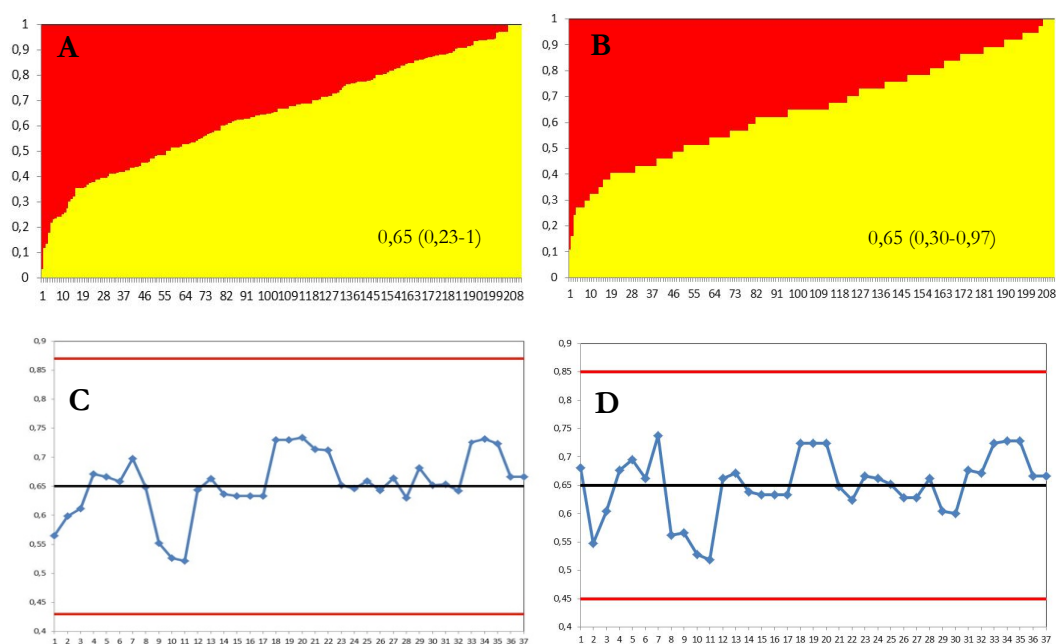


Figura 8: **A e B**, percentagens de ancestralidade nos cromossomas X da amostra total obtidas com todas as imputações (A) e só com imputações de probabilidade superior a 75% (B). Cada cromossoma é representado por uma linha vertical em que a cor amarela é proporcional à percentagem de ancestralidade africana e a cor vermelha proporcional à percentagem de ancestralidade europeia. Em cada painel apresentam-se as percentagens médias de ancestralidade africana e os correspondentes intervalos de confiança a 95% (entre parêntesis). **C e D**, Percentagens de imputações de ancestralidade africana em cada um dos 37 SNPs do cromossoma X. A linha preta representa as proporção médias de ancestralidade africana na amostra total obtida com todos os SNPs. As linhas vermelhas representam os limites de 1 desvio padrão desta média. No painel C usaram-se todas as imputações. No painel D usaram-se apenas imputações com probabilidade superior a 75%.

Com os dados disponíveis é ainda possível calcular as médias das contribuições africanas e europeias para cada um dos 37 marcadores nos 210 indivíduos que compõem a amostra (Figura 8C e D). Esta análise permite verificar se ao longo do cromossoma há regiões genómicas em que as ancestralidades africana ou europeia estão sistematicamente mais representadas do que a média, o que pode sugerir a presença de loci seleccionados nessas regiões [1, 26, 27]. Também neste caso os resultados obtidos com os dois critérios de imputação são muito semelhantes. Nunca se verificaram desvios às proporções médias de ancestralidade que ultrapassassem 1 desvio padrão, o que sugere que as oscilações em torno da média se devem ao acaso (Figura 8C e D).

3.1.2.2 Comparação da miscigenação individual nas diferentes ilhas

Na figura 9, mostram-se as distribuições dos valores de miscigenação individual nas seis ilhas de Cabo Verde que compõem a nossa amostra. Na ilha de Santiago a média da contribuição de origem africana (81%) é claramente mais elevada do que nas outras ilhas (59%-68%). Todas as distribuições de miscigenação individual são significativamente diferentes da de Santiago com o teste de Kolmogorov-Smirnov de duas amostras para um limite de significância de 0,05 (Figura 10). Quando se aplica a correcção de Bonferroni para testes múltiplos, o nível de significância cai para 0,003 e apenas as comparações entre Santiago e Fogo e Santiago e São Nicolau se mantêm significativas (Figura 10). Nenhuma

das comparações entre as ilhas de Santo Antão, São Vicente, São Nicolau, Fogo e Boavista deu resultados significativos (Figura 10).

Este padrão é no geral concordante com a história do povoamento do arquipélago discutida na introdução. Santiago foi a primeira ilha a ser povoada e a que recebeu mais escravos. É natural, por isso, que a razão demográfica entre escravos africanos e colonos europeus fosse maior em Santiago do que noutras ilhas. São Nicolau e Santo Antão são “ilhas camponesas”, povoadas a partir do século XVII por homens livres do Fogo e Santiago, provavelmente com níveis miscigenação elevados [7]. São Vicente foi povoada numa fase mais tardia com habitantes da ilha vizinha de Santo Antão (Figura 1). A colonização da Boavista terá sido idêntica à de São Vicente, mas com imigrantes de São Nicolau (Figura 1). A maior contribuição europeia observada nas populações de São Nicolau, Santo Antão, São Vicente e Boavista, reflecte assim as características do segundo ciclo de povoamento do arquipélago e das fases que se seguiram, nas quais a relevância demográfica da população mestiça se tornou cada vez mais vincada [7].

A única discordância é a diferença entre os níveis de miscigenação de Santiago e do Fogo (Figuras 9 e 10). Como foi referido, nos séculos XV e XVI as duas ilhas foram os principais centros populacionais de Cabo Verde, tendo mantido sistemas de plantação semelhantes no quadro de sociedades escravocratas [6]. Nestas sociedades, os escravos rurais eram predominantes e os escravos domésticos, associados a maior miscigenação, uma minoria. Como se explica então que, apesar da partilha de sistemas sócio-económicos, o Fogo tenha uma contribuição europeia significativamente mais elevada que Santiago? É provável que a resposta esteja nas diferenças de sucesso reprodutivo das comunidades de “escravos rurais” das duas ilhas.

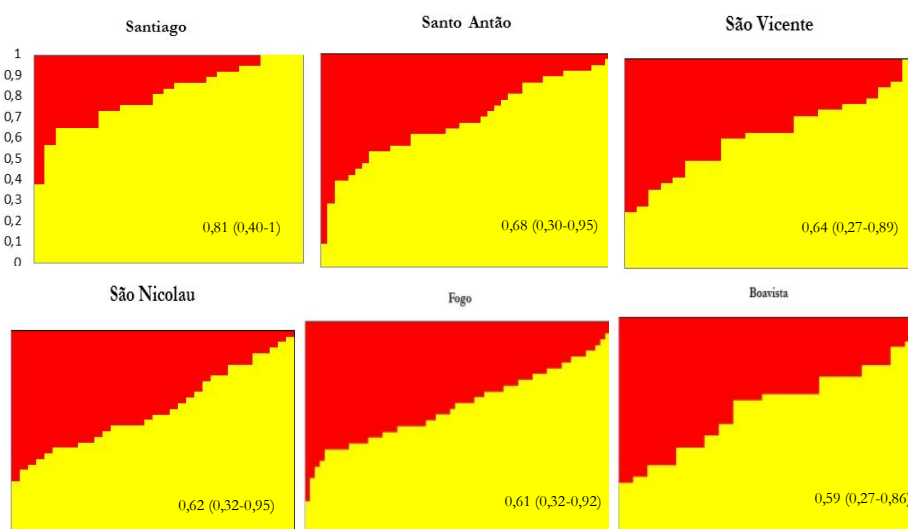


Figura 9: Percentagens de ancestralidade nos cromossomas X de diferentes ilhas de Cabo Verde. A representação é semelhante à das figuras 8A e 8B. Em cada painel apresentam-se as percentagens médias de ancestralidade africana e os correspondentes intervalos de confiança a 95% (entre parêntesis).

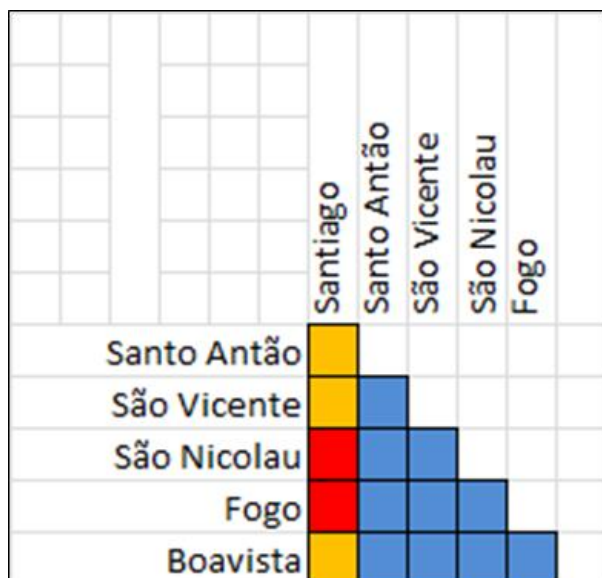


Figura 10: Significância dos valores do teste de Kolmogorov-Smirnov para duas amostras usado para comparar as distribuições da percentagem de ancestralidade africana nas diferentes ilhas de Cabo Verde. Os valores não significativos estão representados a azul; os valores significativos só para o limite de 0,05 estão representados a laranja; os valores significativos também para o limite de 0,003 (obtido pela correcção de Bonferroni) estão representados a vermelho.

Uma das principais características dos sistemas de plantação era a incapacidade das populações de escravos se sustentarem, devido à mortalidade muito elevada [11]. Só a constante importação de escravos compensava a mortalidade e mantinha a disponibilidade de mão-de-obra. Com o declínio das sociedades escravocratas, o predomínio demográfico das populações de origem africana só se reflectiria de modo efectivo nas gerações posteriores se as populações de escravos, e seus descendentes livres, pudessem aumentar progressivamente a sua capacidade reprodutiva. De contrário, a interrupção mais ou menos súbita da importação de escravos levaria ao apagamento dos vestígios da presença africana. Em Santiago, a capacidade de sobrevivência de comunidades de escravos fugitivos e a transição progressiva do regime escravocrata para uma sociedade camponesa a partir do século XVIII [6] terão garantido o sucesso reprodutivo das populações de origem africana, reflectido nos níveis mais baixos de contribuição europeia que hoje se observam na ilha. Pelo contrário, no Fogo a estrutura escravocrata manteve-se inalterada até bastante mais tarde [6]. É possível que no Fogo, as elevadas mortalidades entre os escravos rurais e as dificuldades de sobrevivência de grupos de escravos *fujões* tenham diminuído muito a capacidade reprodutiva das comunidades africanas, que terão acabado por ficar sub-representadas no património genético da ilha. Apenas os mestiços descendentes de escravas domésticas, alguns deles herdeiros dos donos das terras, teriam tido suficiente capacidade reprodutiva para influenciar os padrões de miscigenação observados no Fogo. Deste modo, o que poderia ser inicialmente tomado como um sinal de tolerância inter-racial, acaba por revelar a natureza repressiva do último bastião escravocrata do arquipélago.

3.1.2.3 Comparação da miscigenação em diferentes compartimentos do genoma

A comparação das estimativas de miscigenação obtidas com marcadores de diferentes compartimentos do genoma é útil para avaliar a influência relativa dos sexos feminino e masculino nos padrões de diversidade genética.

Como foi dito na introdução, o projecto em que este trabalho está inserido envolve o estudo de vários marcadores autossómicos e do cromossoma Y nas populações de Cabo Verde. Por outro lado, há trabalhos anteriores em que foi publicada informação sobre a variação genética do DNA mitocondrial (mtDNA) no arquipélago [28]. De acordo com estes trabalhos, a contribuição parental feminina para o arquipélago é quase toda de origem africana [28], o que implica que a miscigenação de Cabo Verde resultou essencialmente de uniões entre homens europeus e mulheres africanas, tal como seria de esperar da evidência histórica [10].

Esta assimetria nos padrões de cruzamento faz com que as percentagens de contribuição africana e europeia não sejam as mesmas em todos os compartimentos do genoma. Na figura 11 mostra-se a variação da contribuição africana no cromossoma X e nos autossomas para diferentes contribuições masculinas medidas pelo cromossoma Y, assumindo que a contribuição feminina é integralmente africana e que a miscigenação entre populações parentais ocorreu numa única geração, tendo-se mantido condições de panmixia nas gerações seguintes. Para contribuições africanas masculinas inferiores a 100% (ausência de miscigenação), as proporções de ancestralidade africana no cromossoma X são sempre maiores que nos autossomas que, por sua vez, são maiores que as do cromossoma Y (Figura 11). Estas diferenças reflectem o número relativo de cromossomas envolvidos nos cruzamentos que originam descendentes miscigenados. Quando homens europeus se cruzam com mulheres africanas, em cada casal o cromossoma Y será de origem europeia, enquanto 50% dos autossomas e 67% ($2/3$) dos cromossomas X serão de origem africana.

Na tabela 3 comparam-se as contribuições médias africanas no cromossoma X com as do cromossoma Y e dos autossomas, obtidas no âmbito do projecto de caracterização genética do arquipélago.

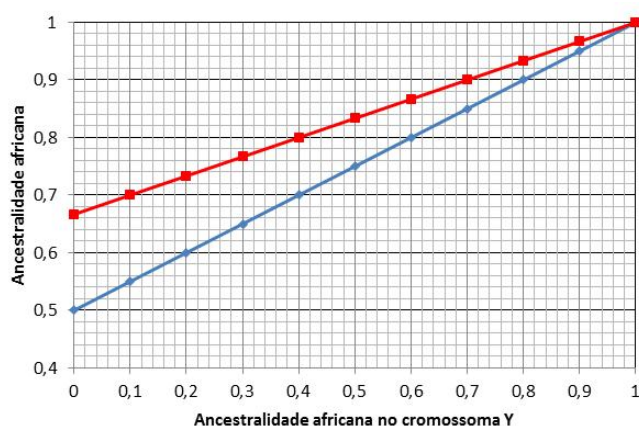


Figura 11: Variação na proporção da ancestralidade africana em autossomas (linha azul) e no cromossoma X (linha vermelha) para diferentes valores de contribuição africana do cromossoma Y. Os cálculos foram feitos no pressuposto de que a contribuição feminina para a população miscigenada é toda de origem africana.

Tabela 3: Proporções médias de ancestralidade africana em seis ilhas de Cabo Verde calculadas com marcadores do cromossoma Y, do cromossoma X e dos autossomas.

	Ancestralidade africana				
	Cromossoma Y ^a	Autossomas ^b (Obs)	Autossomas ^c (Esp)	Cromossoma X ^d (Obs)	Cromossoma X ^e (Esp)
Santo Antão	0,36	0,56	0,68	0,68	0,79
São Vicente	0,21	0,56	0,61	0,64	0,74
São Nicolau	0,1	0,56	0,55	0,62	0,70
Boavista	0,05	0,59	0,53	0,59	0,68
Fogo	0,09	0,53	0,55	0,61	0,70
Santiago	0,57	0,65	0,79	0,81	0,86
Total	0,27	0,57	0,64	0,65	0,76

a. Com base em 10 SNPs informativos.

b. Com base em 50 AIMs autossômicos (Obs=observado).

c. Valor esperado (Esp) com base nas proporções de ancestralidade observadas no cromossoma Y (ver Figura 11).

d. Com base no presente trabalho

e. Valor esperado com base nas proporções de ancestralidade observadas no cromossoma Y (ver Figura 11).

Os resultados obtidos para os diferentes compartimentos do genoma confirmam que a ilha de Santiago é a que tem maior contribuição africana (Tabela 3). Como seria de esperar numa situação em que os dois sexos não participam de igual forma na miscigenação (Figura 11), a ancestralidade africana é significativamente maior nos autossomas e no cromossoma X do que no cromossoma Y ($P < 0,05$, em ambos os casos, usando o teste de Wilcoxon para amostras emparelhadas). No entanto, os valores calculados para os autossomas são ainda assim mais baixos do que seria de esperar com o modelo de miscigenação utilizado na figura 11 ($P < 0,05$ com o teste de Wilcoxon para amostras emparelhadas). O mesmo acontece no cromossoma X. Por exemplo, se a miscigenação tivesse ocorrido numa única geração, em condições de panmixia esperar-se-ia que à proporção de 36% de cromossomas Y com origem africana de Santo Antão correspondesse um valor de ancestralidade de 79% no cromossoma X e não 68%, como foi observado (Tabela 3). Estas discrepâncias mostram que o modelo de miscigenação utilizado para calcular os valores esperados não descreve adequadamente a estrutura da população cabo-verdiana. No caso de Santo Antão, um efeito de fundador associado a um colono mestiço com um cromossoma Y de origem africana, por exemplo, poderia causar um excesso de ancestralidade africana do cromossoma Y relativamente ao cromossoma X do tipo verificado na tabela 3. O número de cenários de violação da panmixia que explicam este tipo de desvios é, contudo, muito elevado e serão necessários mais estudos para explorar as várias explicações alternativas.

3.1.2.4 Miscigenação e recombinação

A transição de ancestralidade em alelos de marcadores consecutivos (suturas) corresponde a pelo menos uma recombinação entre cromossomas de origem africana e europeia no segmento delimitado por esses marcadores. Por isso, é de esperar que as suturas sejam mais frequentes nos segmentos com maior fracção de recombinação, caso as transições de ancestralidade resultem de recorrência da recombinação.

Em Cabo Verde, confirma-se que a percentagem de suturas está fortemente correlacionada ($R^2 \sim 0,80$) com as estimativas de fracção de recombinação obtidas por outros métodos - neste caso as disponibilizadas pelo projecto HapMap (Figura 12; cf. secção 2.3 e Tabela 1). Esta concordância mostra que a população do arquipélago pode ser usada para gerar mapas de recombinação baseados na distribuição de suturas, que poderão ser depois comparados com os mapas obtidos com outras metodologias [17, 23].

A figura 13 mostra que a recombinação medida em cM por megabase (cM/Mb) não está uniformemente distribuída ao longo das 151 Mb cobertas pelos 37 SNPs do cromossoma X (cf. 2.3 e Tabela 1). Apesar de a média estimada para este cromossoma (1,1cM/Mb) ser próxima da média genómica (1 cM/Mb, [29]), há oscilações consideráveis (0,075-3,21 cM/Mb), com tendência para os valores mais altos se situarem próximo do telómero e para os valores mais baixos se situarem próximo do centrómero. Na região pericentromérica, os segmentos delimitados pelos SNPs rs7061449, rs7886355, rs35972027, rs699863 e rs492933 (loci nº 13 a 17, ver Figura 3) formam a maior região contígua do mapa físico (13,6 Mb) em que a fracção de recombinação média (0,28 cM/Mb) é claramente mais baixa que a média (Figura 13). Esta região contém os marcadores rs7886355, rs35972027, rs699863 e rs492933 (14 a 17), associados muito pouca incerteza na imputação da ancestralidade (Figura 6A e secção 3.1.1), e dentro dela nunca se observou qualquer sutura no segmento que inclui os marcadores rs35972027, rs699863 e rs492933 (15, 16 e 17). A frequência de segmentos de origem africana definidos por estes marcadores na amostra total é semelhante à contribuição média africana calculada com todos os marcadores (0,63 vs. 0,65). Estas observações sugerem que, em cada cromossoma X, a região que contém os marcadores 15 a 17 pode ser um fragmento do cromossoma original, de origem africana ou europeia, que se manteve intacto devido à ausência de recombinação. Se assim for, será possível não só determinar as proporções actuais de ancestralidade africana e europeia de um cromossoma (Figuras 8 e 9), mas também a população ancestral de onde esse cromossoma proveio no momento em que chegou ao arquipélago.

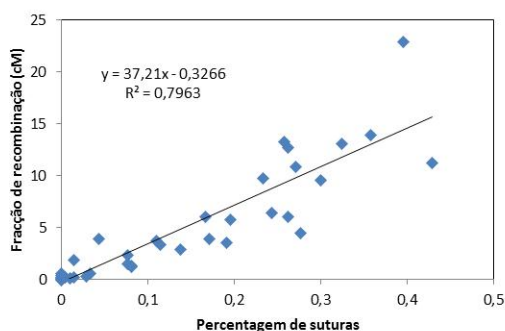


Figura 12: Relação entre a percentagem de transições de ancestralidade (suturas) entre pares de SNPs consecutivos e a fracção de recombinação.

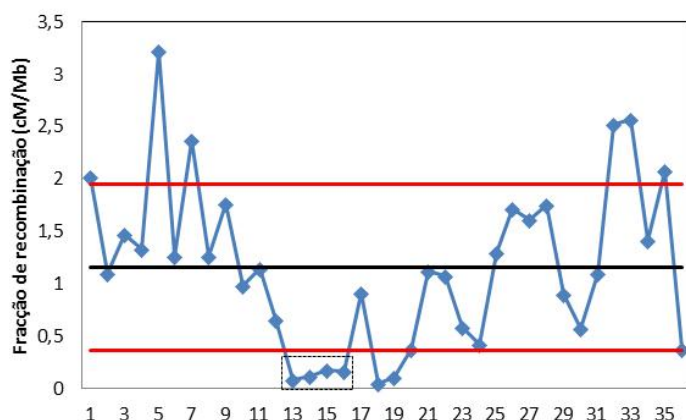


Figura 13: Variação na fracção de recombinação (cM/Mb) entre pares de SNPs consecutivos. A linha a preto mostra a fracção de recombinação média no cromossoma X (1,1 cM/Mb). As linhas vermelhas mostram os limites de variação definidos com 1 desvio padrão. O rectângulo delimita os intervalos entre SNPs situados na região contígua justacentromérica com menor fracção de recombinação.

Com esta informação adicional, é possível analisar separadamente os efeitos da miscigenação em cromossomas com origem africana e europeia. Por exemplo, na Figura 14 mostram-se as percentagens de ancestralidade africana em cromossomas cuja proveniência africana ou europeia foi estipulada a partir do fragmento que contém os marcadores 15 a 17. Os cromossomas com origem em África representam 63% da amostra total de cromossomas X e retêm uma percentagem média de ancestralidade africana de 74% (com um intervalo de variação entre 32% e 100%). Os cromossomas com origem europeia representam 37% da amostra total e a sua percentagem média de ancestralidade africana é de 49% (com um intervalo entre 11% e 89%). Note-se como nos “cromossomas europeus”, em menor número, o intervalo de variação tem maior amplitude do que nos “cromossomas africanos”. Note-se, também como os valores médios de ancestralidade africana em cada um dos conjuntos de cromossomas (74% e 49%) ainda diferem do valor médio total (63%). Esta diferença indica que o processo de amalgamento total dos cromossomas ainda não está concluído. Com efeito, de acordo com o modelo de miscigenação descrito na secção 3.1.2.3, quando as populações europeias e africanas contactaram haveria 63% de cromossomas X africanos e 37% de cromossomas X europeus. Nessa altura, porém, a percentagem de ancestralidade africana seria de 100% nos cromossomas de origem africana e de 0% nos cromossomas de origem europeia. Esta situação pode ser descrita com *mistura de cromossomas*. Quando a situação de total amalgamento for atingida, haverá *cromossomas misturados*, ou seja, a percentagem de ancestralidade africana será igual ao valor médio de 63%, quer nos cromossomas X inicialmente europeus, quer nos cromossomas X inicialmente africanos.

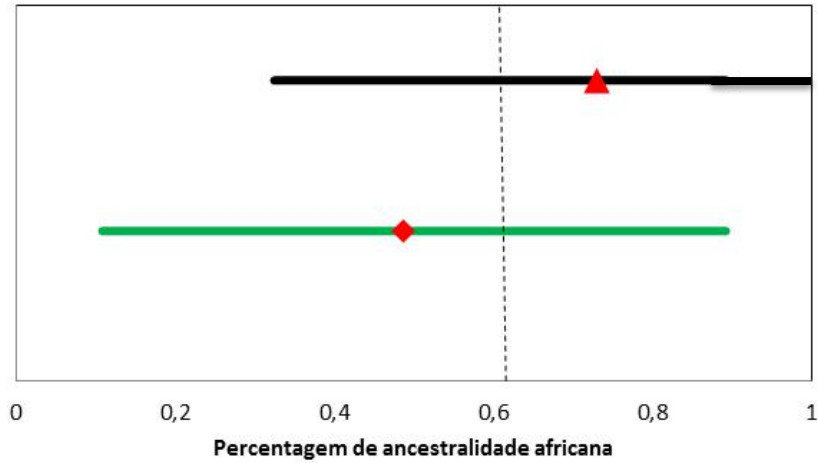


Figura 14: Intervalos de variação da percentagem de ancestralidade africana em cromossomas originalmente provenientes de África (a preto) e da Europa (a verde). Os valores médios de ancestralidade africana estão representados por triângulos (cromossomas africanos) e losangos (cromossomas europeus) vermelhos. A linha vertical tracejada indica a percentagem média de ancestralidade africana, independentemente da proveniência dos cromossomas.

3.1.2.5 Tempo decorrido desde o início da miscigenação

A noção de que é possível identificar a origem de um cromossoma miscigenado pode ser usada para determinar o tempo que decorreu desde o início da miscigenação.

Assumindo que há uma região cromossómica que preserva a ancestralidade, é possível quantificar a proporção de cromossomas que conserva a ancestralidade original à medida que nos afastamos dessa região. Como essa proporção depende da fracção de recombinação e do tempo que decorreu desde que cromossomas com diferenças origens diferentes iniciaram o contacto, é possível calcular há quanto se iniciou a miscigenação se se souber qual é a fracção de recombinação.

Para aplicar este princípio, calculou-se a proporção de cromossomas que mantém a ancestralidade original usando SNPs situados a diferentes distâncias da região que abarca os marcadores rs35972027, rs699863 e rs492933 (loci 15, 16 e 17). A figura 15 mostra os resultados desta contagem para os 77 cromossomas com origem europeia (Figura 15 A) e para os 133 cromossomas de origem africana (Figura 15 B). Note-se, como há maior retenção da origem em cromossomas africanos, mais numerosos, do que em cromossomas europeus, menos numerosos e, portanto, mais frequentemente emparelhados com cromossomas de origem diferente.

Para cada SNP, e para cada distância, o número de gerações decorrido desde a miscigenação (G) foi calculado de acordo com a metodologia descrita por Stephens et al. [30] e Reich et al. [31] para datar mutações, através de:

$$G = 1,5 \cdot (-\ln P / c)$$

em que P é a fracção de cromossomas que, para determinada distância, mantém a ancestralidade original (figura 15) e c é a taxa *efectiva* de recombinação. Acrescenta-se o factor 1.5 porque a recombinação no cromossoma X só ocorre no sexo feminino (2/3 do tempo), o que corresponde a um aumento de 3/2 do número de gerações relativamente aos autossomas.

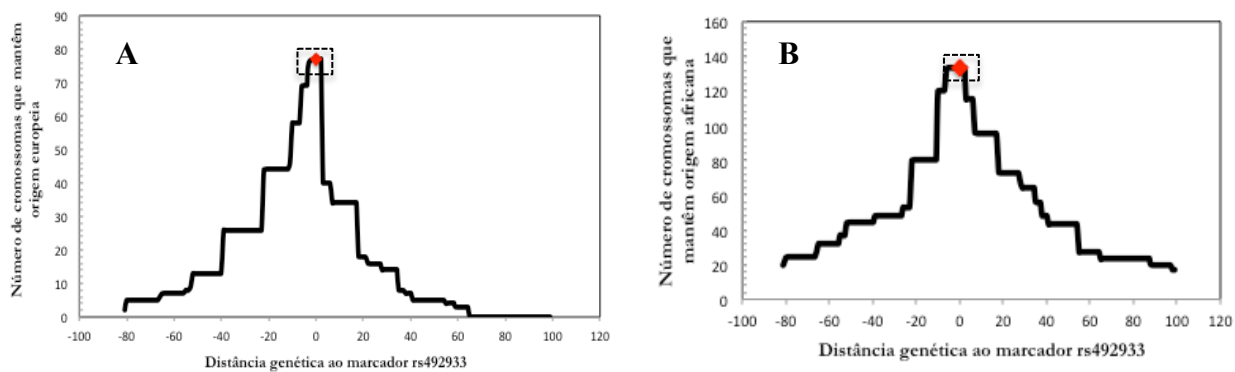


Figura 15: Número de cromossomas de origem europeia (A) e africana (B) que mantêm a ancestralidade original em SNPs situados a distâncias (cM) crescentes da região justacentromérica identificada com um ponto vermelho e um rectângulo. Ver detalhes no texto.

A taxa efectiva de recombinação corresponde à fracção de entrecruzamentos meióticos que produzem, de facto, uma alteração da ancestralidade de um fragmento cromossómico. Por exemplo, numa situação em que haja cerca de 63% de cromossomas de origem africana (como é se estima que seja o caso de Cabo Verde) só 63% das recombinações que envolvem um cromossoma europeu é que modificam a ancestralidade do fragmento trocado nessas recombinações. Os restantes 37% envolvem recombinações entre cromossomas exclusivamente europeus e não alteram a ancestralidade. Na verdade, o problema é complicado pelo facto de os cromossomas de origem africana também adquirirem fragmentos europeus à medida que o tempo passa. Nestes casos, há a possibilidade de uma recombinação entre cromossomas europeus e africanos não modificar, ou mesmo regenerar a origem ancestral. Uma solução é calcular a taxa efectiva de recombinação para cromossomas europeus tendo em conta as configurações cromossómicas actualmente observadas nos cromossomas africanos e vice-versa. Por exemplo, se a fracção de recombinação entre dois SNPs for 0.1 e houver 50% de cromossomas africanos que adquiriram segmentos de origem europeia a partir do SNP mais distal, a probabilidade de se observar uma alteração de ancestralidade num cromossoma europeu será: $0.1 \times 0.5 \times 0.63$. Na prática esta correcção é exagerada, porque no início da história das populações miscigenadas todas as recombinações entre cromossomas de origem diferente produziam alterações de ancestralidade. Por esta razão, optou-se por apresentar estimativas obtidas com uma correcção mínima – correspondente no exemplo a 0.63×0.1 - e com uma correcção máxima - correspondente no exemplo a $0.1 \times 0.5 \times 0.63$. À correcção mínima correspondem estimativas de t mais elevadas e tempos mais curtos. À correcção máxima correspondem estimativas de t mais baixas e tempos mais recuados.

Na Tabela 4 mostram-se as estimativas do tempo decorrido desde o início da miscigenação obtidas com cromossomas de origem europeia e africana. Calculou-se uma data para a distância entre cada marcador e o SNP rs492933 localizado no segmento mais preservado. Apresentam-se separadamente as estimativas obtidas para cada lado da região preservada, usando um tempo de geração de 28 anos [32]. As estimativas pontuais oscilam entre ~ 300 e ~ 500 anos. A média das estimativas máxima e mínima obtida com cromossomas de origem europeia é de 450 anos é muito próxima dos 500 anos que decorreram desde o povoamento de Cabo Verde, indicando que a miscigenação se iniciou logo na primeira fase do povoamento

Tabela 4: Estimativas do tempo (em anos) decorrido desde o início da miscigenação entre africanos e europeus em Cabo Verde.

	Baseada em “cromossomas africanos” ¹		“Baseada em cromossomas europeus” ²	
	Estimativa máxima ³	Estimativa mínima ⁴	Estimativa Máxima	Estimativa mínima
Marcadores à esquerda ⁶	405 (316-529) ⁵	386 (299-503)	422 (284-683)	330 (208-615)
Marcadores à direita ⁷	285 (215-376)	277 (208-362)	543 (397-772)	494 (361-684)
Média	345 (266-453)	332 (254-433)	483 (340-728)	412 (285-650)

¹ Cromossomas com origem africana inferida na região preservada com os marcadores rs35972027, rs699863 e rs492933.

² Cromossomas com origem europeia inferida na região preservada.

³ Com taxa efectiva de recombinação calculada através da distribuição de segmentos de ancestralidade actual.

⁴ Com taxa efectiva de recombinação calculada apenas com as frequências de cromossomas europeus e africanos.

⁵ Intervalos de confiança de 95% baseados no desvio-padrão de P (ver fórmula no texto).

⁶ Marcadores localizados na direcção do braço Xp relativamente à região preservada.

⁷ Marcadores localizados na direcção do braço Xq relativamente à região preservada.

A média equivalente obtida com cromossomas de origem Africana (~340 anos) é mais recente em cerca de um século. É importante notar que nas estimativas da Tabela 4 se assumiu um modelo simplificado de panmixia após um único pulso de miscigenação. Em caso de violação das condições deste modelo o tempo decorrido desde a miscigenação tende a ser subestimado [17, 22, 23]. Por exemplo se não houver panmixia e uma fracção dos habitantes preferir cruzar-se com pessoas com a mesma origem, haverá um efeito de retardamento da miscigenação que não é tido em conta nos cálculos apresentados na Tabela 4. É também provável que este enviesamento se possa fazer sentir com maior intensidade nos cromossomas de origem africana, que tendem a produzir estimativas mais recentes do que os cromossomas de origem europeia. De qualquer modo, as datações obtidas estão, ainda assim, razoavelmente de acordo com os dados históricos sobre o povoamento de Cabo Verde.

3. 2. Resultados obtidos com STRs e SNPs

Como se referiu na *secção 3.1*, a percentagem de situações em que não foi possível deduzir a ancestralidade de um alelo com uma probabilidade superior a 75% - correspondente a probabilidades de ancestralidade ente 25% e 75% - pode ser usada para aferir a incerteza associada à imputação de ancestralidade. Com os 37 SNPs, essa percentagem foi de 12,5%. Quando se usaram apenas 14 microssatélites a percentagem foi, como também se referiu, muito superior (86%). Quando se usam simultaneamente os SNPs e 14 STRs (num total de 51 marcadores) a incerteza é de 18,2%, pouco maior do que a correspondente aos SNPs. Estes resultados indicam que a ambiguidade na estimativa da ancestralidade dos alelos dos STRs diminui fortemente quando a informação disponível é ancorada em SNPs cujos alelos têm grandes diferenças de frequência na Europa e em África.

A utilização conjunta de SNPs e STRs aumenta a densidade de marcadores e permite reavaliar algumas das estimativas obtidas anteriormente apenas com SNPs. Tal como se fez na análise de SNPs, mostram-se apenas os resultados obtidos com o critério de imputação mais relaxado, em que se considera que um alelo tem proveniência africana ou europeia quando a sua probabilidade de origem é superior a 50% (cf. *secção 3.2.1.4*). Os resultados obtidos com este critério - que inclui uma percentagem de 12,5% de imputações com probabilidades entre 25% e 75% - praticamente não diferem das estimativas obtidas a partir de imputações com probabilidades superiores a 75% (não mostrado).

Na Tabela 5 comparam-se as proporções médias de ancestralidade africana obtidas só com SNPs e com SNPs e STRs. Os termos de comparação são as médias das ancestralidades individuais em todos os cromossomas X analisados (cf. Figura 9 e Tabela 3) e as médias das ancestralidades individuais obtidas em cromossomas X com origem africana e europeia, de acordo com a abordagem discutida na *secção 3.2.1.4*. (cf. Figura 14). A semelhança entre as estimativas obtidas com diferentes densidades de marcadores é notável e sugere que os resultados obtidos só com SNPs são suficientemente robustos. Esta semelhança estende-se a outros aspectos da análise de miscigenação, incluindo os perfis de ancestralidade (ver Figura 8), a distribuição dos pontos de recombinação (ver Figura 12) e as datações do início da miscigenação. Para evitar redundância, essas análises não são aqui apresentadas.

Tabela 5: Proporções médias de ancestralidade africana em seis ilhas de Cabo Verde calculadas com marcadores do cromossoma X

	Ancestralidade africana					
	Em todos os cromossomas		Em cromossomas Europeus		Em cromossomas Africanos	
	SNPs	SNPs+STRs	SNPs	SNPs+STRs	SNPs	SNPs+STRs
Santo Antão	0,68	0,67	0.54	0,58	0.77	0,76
São Vicente	0,64	0,63	0.41	0,40	0.71	0,71
São Nicolau	0,62	0,61	0.44	0,45	0.71	0,70
Boavista	0,59	0,60	0.49	0,52	0.69	0,68
Fogo	0,61	0,62	0.48	0,5	0.70	0,71
Santiago	0,81	0,82	0.59	0,59	0.87	0,90
Total	0,65	0,66	0.49	0,51	0.74	0,74

Há, contudo, uma análise que pode beneficiar da maior densidade de marcadores obtida com STRs e SNPs, em que se explora a distribuição do tamanho de fragmentos contínuos em que a ancestralidade se mantém constante [20, 22]. Para simplificar a análise, optou-se por estudar apenas a distribuição do tamanho dos fragmentos que abrangem a região pericentromérica que permite identificar a proveniência africana ou europeia de cada cromossoma (cf. *secções 3.2.1.4 e 3.1.2.5*). Esta análise é, de certo modo, complementar da abordagem seguida na Figura 15.

Na Figura 16 A mostra-se uma representação esquemática do tamanho relativo de fragmentos com ancestralidade europeia e africana em cromossomas de proveniência europeia e africana, respectivamente. Note-se como a maior tendência para os cromossomas de origem africana (a amarelo) preservarem a sua ancestralidade (cf. *secções 3.2.1.4 e 3.1.2.5*) se traduz na retenção de fragmentos com maior tamanho relativo em torno da região pericentromérica. É também interessante verificar como muitos cromossomas (sobretudo de origem europeia) partilham o mesmo ponto de interrupção da continuidade dos fragmentos, no braço Xq. Este ponto coincide com o primeiro pico de elevação da fracção de recombinação na vizinhança da região pericentromérica (Figura 16 B).

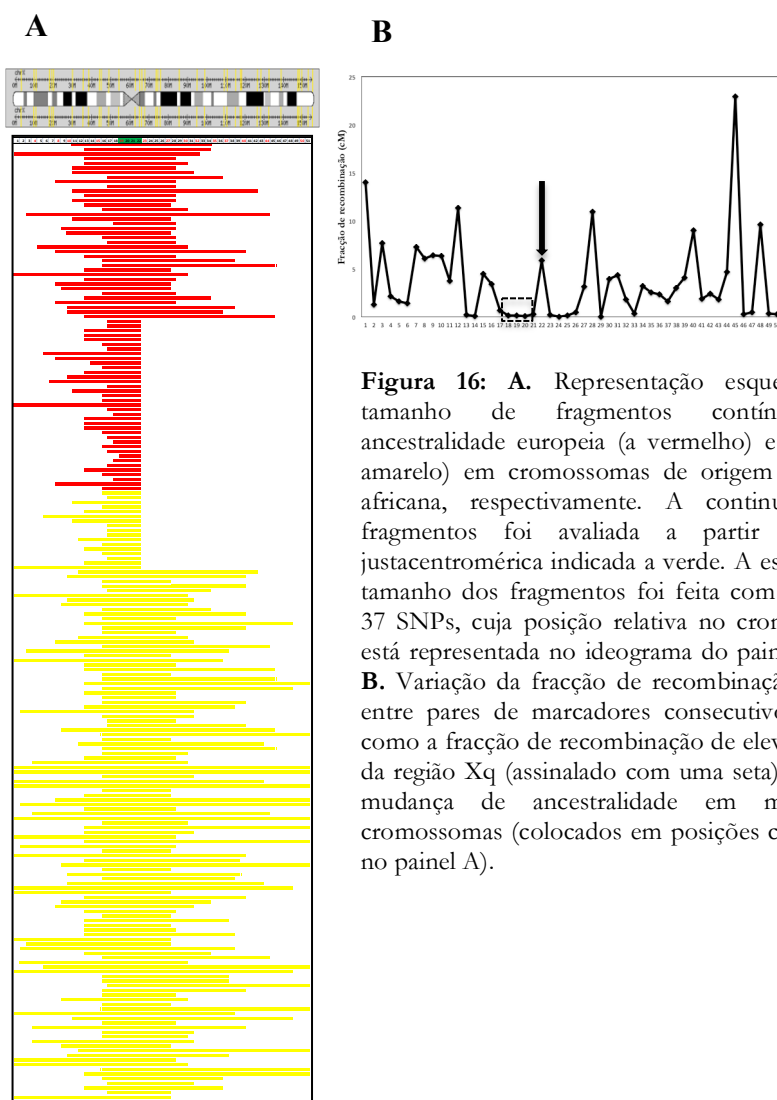


Figura 16: **A.** Representação esquemática do tamanho de fragmentos contínuos com ancestralidade europeia (a vermelho) e africana (a amarelo) em cromossomas de origem europeia e africana, respectivamente. A continuidade dos fragmentos foi avaliada a partir da região justacentromérica indicada a verde. A estimativa do tamanho dos fragmentos foi feita com 14 STRs e 37 SNPs, cuja posição relativa no cromossoma X está representada no ideograma do painel superior. **B.** Variação da fracção de recombinação (em cM) entre pares de marcadores consecutivos. Note-se como a fracção de recombinação de eleva no ponto da região Xq (assinalado com uma seta) em que há mudança de ancestralidade em muitos dos cromossomas (colocados em posições consecutivas no painel A).

O tamanho dos fragmentos cromossômicos que incluem a região pericentromérica pode ser usado para avaliar a dinâmica do processo de miscigenação. Se este processo estiver completo e a população amalgamada, a contribuição das populações parentais para cada cromossoma não estará correlacionada com o tamanho dos fragmentos. De facto, quando as condições de equilíbrio são alcançadas, espera-se que os fragmentos contínuos que preservam uma origem africana ou europeia tenham um tamanho constante [19, 20]. Nessas condições, a maior ou menor proporção de ancestralidade europeia ou africana dependerá do número de fragmentos com uma ou outra ancestralidade, mas espera-se que esses fragmentos estejam alternados. Pelo contrário, quando o processo de miscigenação está incompleto - por ser recente ou por ser retardado na ausência de panmixia - haverá tendência para correlação entre a proporção de ancestralidade de um cromossoma e o tamanho de fragmentos contínuos que conservam essa ancestralidade [19, 20]. Na Figura 17 apresentam-se os resultados de análises de regressão de Spearman e Pearson em que se procura avaliar até que ponto a ancestralidade de um cromossoma está correlacionada com o tamanho dos fragmentos representados na Figura 16. Nos cromossomas europeus o coeficiente de correlação de Spearman é de 0,60 e o de Pearson é de 0,68 e o (Figura 17 A e B, respectivamente). Nos cromossomas africanos o coeficiente de correlação de Spearman e de Pearson são ambos de 0,72 (Figura 17 C e D). Estas correlações mostram que o processo de miscigenação ainda não atingiu o equilíbrio, em acordo com resultados de outras análises (cf. *secção 3.2.1.4*). Note-se como a intensidade da correlação é maior nos cromossomas de origem africana do que nos cromossomas de origem europeia, que tendem a apresentar um maior nível de mosaicism, aproximando-se mais rapidamente do equilíbrio, devido à sua maior taxa efectiva de recombinação (cf. *secção 3.1.2.5*). A disparidade nos níveis de mosaicism reflecte-se ainda no tamanho médio dos fragmentos que retêm a ancestralidade original, incluindo a região pericentromérica: 29,45 cM nos cromossomas europeus e 62,61 cM nos cromossomas africanos. Curiosamente, a razão entre estes dois valores ($62,61/29,45=2,13$) é próxima da razão entre a ancestralidade média africana e europeia ($0,66/0,34=1,94$).

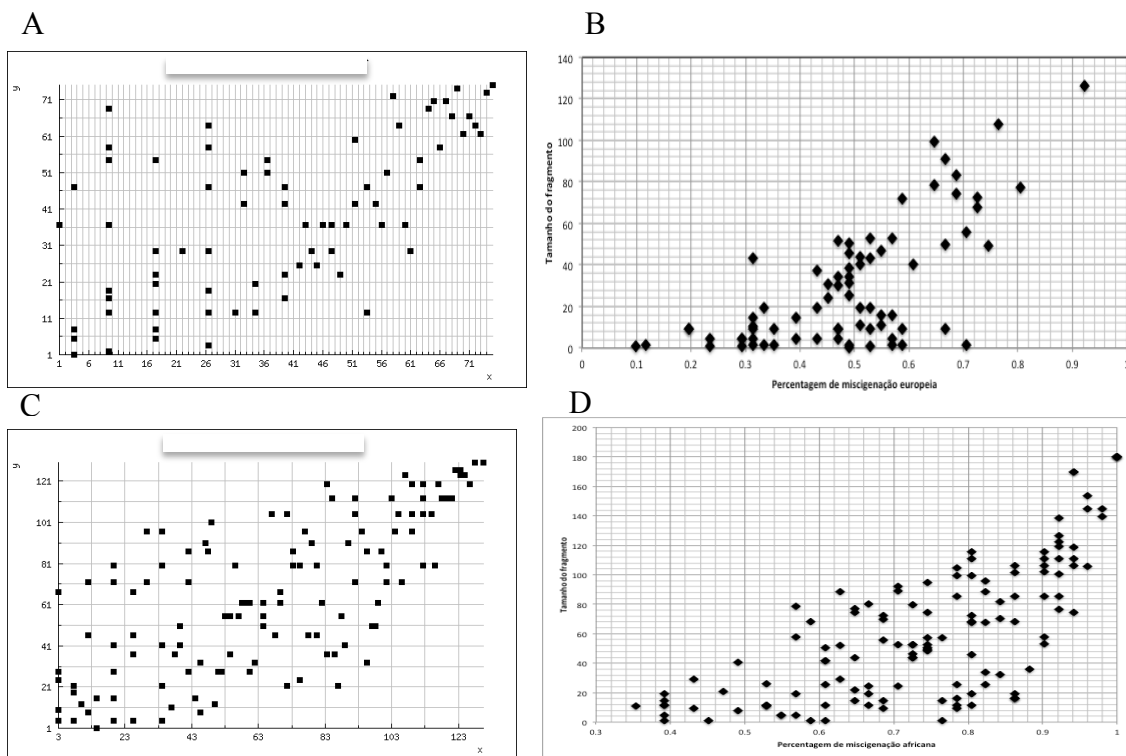


Figura 17: A e C, Relação entre a ordem do tamanho de fragmentos contínuos (eixo dos Ys) e a ordem da percentagem de ancestralidade original (eixo dos Xs) em cromossomas X de origem europeia (A) e africana (C). B e D, Relação entre o tamanho de fragmentos contínuos (eixo dos Ys) e a percentagem de ancestralidade original (eixo dos Xs) em cromossomas de origem europeia (B) e africana (D).

A verificação de maior mosaicismismo em cromossomas de origem europeia pode ser usada para aumentar a eficácia de estudos de mapeamento por miscigenação que envolvam *loci* do cromossoma X. Neste estudos, a delimitação de genes responsáveis por características com grande diferença de prevalência em África e na Europa beneficia da presença de segmentos com ancestralidade europeia e africana relativamente curtos. Por isso, a utilização de cromossomas com reconhecida origem europeia pode ser mais útil do que o estudo conjunto de todos os cromossomas, independentemente da sua origem. Neste caso, a inclusão de cromossomas africanos aumentaria o número de fragmentos de grandes dimensões que acabariam por prejudicar a interpretação dos resultados.

4. Conclusões

O estudo-piloto agora realizado aumentou consideravelmente o nível de detalhe da análise da miscigenação do arquipélago. A abordagem adoptada, pela sua relativa simplicidade, deve ser vista como um primeiro passo que deverá ser estendido a outros cromossomas com uma densidade de marcadores mais elevada, tal como já tem sido feito, por exemplo, em populações afro-americanas (REF). De qualquer modo, com as condições experimentais adoptadas foi já possível obter uma *prova de princípio* e revelar aspectos importantes da dinâmica da miscigenação em Cabo Verde que até aqui não estavam suficientemente explorados. Entre esses aspectos destacam-se os seguintes:

- i) Utilizando SNPs com grandes diferenças de frequências alélicas entre africanos e europeus (AIMs), foi possível imputar ancestralidades alélicas com um nível de ambiguidade relativamente pequeno (12,5%), que não interferiu na robustez dos resultados apesar da baixa densidade de marcadores (em média 1 marcador por cada 4,86 cM).
- ii) Um número relativamente pequeno de SNPs com características de AIMS pode servir para *ancorar* outros *loci* escolhidos ao acaso (por exemplo STRs) que permitam aumentar a densidade de marcadores ao longo de um cromossoma.
- iii) Os principais factores que influenciam a incerteza das imputações de ancestralidade num marcador são a diferença de frequências alélicas entre populações parentais e a distância genética ao marcador mais próximo.
- iv) Tendo em conta a tendência para a supressão da recombinação nas regiões justacentroméricas, foi possível identificar, na região do centrómero, zonas cromossómicas que preservavam a ancestralidade original de um cromossoma, independentemente dos seus níveis de miscigenação actuais. Esta informação permitiu analisar separadamente os efeitos da miscigenação em cromossomas com origem africana e europeia.
- v) A proporção média de ancestralidade africana nos cromossomas X de Cabo Verde (~65%) indica que a contribuição feminina para o componente africano foi predominante e que a miscigenação resultou de uniões entre mulheres africanas e homens europeus. Esta proporção não está, contudo, distribuída uniformemente no arquipélago. Na ilha de Santiago há uma contribuição africana (81%) significativamente superior à das outras ilhas (59-68%). Em geral, esta diferença reflecte bem o que se conhece dos ciclos de povoamento do arquipélago. No entanto, na ilha do Fogo há uma aparente discrepância entre níveis de contribuição africana relativamente baixos (0,61) e o predomínio documentado de um sistema escravocrata. Esta discrepância indicia que a capacidade de sobrevivência das comunidades de origem africana do Fogo era muito inferior às das comunidades africanas de Santiago.
- vi) A distribuição dos níveis de miscigenação obtidos com o cromossoma X está qualitativamente de acordo com a informação disponível para autossomas e para o cromossoma Y, em particular no que se refere à maior contribuição africana na ilha de Santiago. No entanto, as estimativas de miscigenação obtidas com

diferentes compartimentos do genoma não são concordantes, mesmo tendo em conta os efeitos da assimetria de sexos nas uniões que promoveram a miscigenação. Esta discordância indica que o processo de miscigenação em Cabo Verde não pode ser descrito com um modelo simples de panmixia, sem fluxo génico das populações parentais ao longo das gerações.

- vii) O cálculo do tempo decorrido desde o início da miscigenação (300 a 500 anos) está de acordo com a história conhecida do arquipélago e sugere que o processo de mestiçagem, à semelhança da formação do crioulo cabo-verdiano, se iniciou no primeiro ciclo de povoamento do arquipélago.
- viii) A análise do processo de miscigenação em cromossomas de origem africana e europeia indica que ainda não se atingiu o amalgamento total característico do equilíbrio, por a miscigenação ser recente e/ou por ausência de panmixia.
- ix) Devido à predominância de cromossomas X de origem Africana (65% vs 35%) os cromossomas de origem europeia tem níveis muito mais elevados de mosaicismo e aproximam-se mais rapidamente do equilíbrio. Esta propriedade pode ser explorada para aumentar o poder de estudos de mapeamento por miscigenação que envolvam *loci* situados no cromossoma X.

5. Referências bibliográficas

1. Jobling MA, Hurles ME, Tyler-Smith C (2004) *Human Evolutionary Genetics*. Garland Science, Abington.
2. Dupanloup I, Bertorelle G, Chikhi L, Barbujani G (2004) Estimating the impact of prehistoric admixture on the genome of Europeans. *Mol Biol Evol* **21**: 1361-1372.
3. Coelho M, Sequeira F, Luiselli D, Beleza S, Rocha J (2009) On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evol Biol* **9**: 80.
4. McKeigue PM (2005) Prospects for admixture mapping of complex traits. *Am J Hum Genet* **76**: 1-7.
5. Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, Hauser SL, Smith MW, O'Brien SJ, Altshuler D, Daly MJ, Reich D (2004) Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* **74**: 979-1000.
6. Correia e Silva A (2001) Espaço, ecologia e economia interna. In Albuquerque L, Madeira Santos ME (coordenadores) *História Geral de Cabo Verde*, vol. 1 (2ª edição). pp 179-236. Centro de Estudo de História e Cartografia Antiga, Instituto de Investigação Científica Tropical. Lisboa. Instituto Nacional de Investigação Cultural de Cabo Verde. Praia.
7. Correia e Silva A (2002) Dinâmicas de decomposição e recomposição de espaços e sociedades. In Albuquerque L, Madeira Santos ME (coordenadores) *História Geral de Cabo Verde*, vol. 3. pp 1-66. Centro de Estudo de História e Cartografia Antiga, Instituto de Investigação Científica Tropical. Lisboa. Instituto Nacional de Investigação Cultural de Cabo Verde. Praia.
8. Amaral I (2001) Cabo Verde: introdução geográfica. In Albuquerque L, Madeira Santos ME (coordenadores) *História Geral de Cabo Verde*, vol. 1 (2ª edição). pp 1-22. Centro de Estudo de História e Cartografia Antiga, Instituto de Investigação Científica Tropical. Lisboa. Instituto Nacional de Investigação Cultural de Cabo Verde. Praia.
9. Baleno I (2001) Povoamento e formação da sociedade. In Albuquerque L, Madeira Santos ME (coordenadores) *História Geral de Cabo Verde*, vol. 1 (2ª edição). pp 125-177. Centro de Estudo de História e Cartografia Antiga, Instituto de Investigação Científica Tropical. Lisboa. Instituto Nacional de Investigação Cultural de Cabo Verde. Praia.
10. Carreira A (2000) *Cabo Verde: formação e extinção de uma sociedade escravocrata* (2ª edição). Instituto de Promoção Cultural. Praia.
11. Curtin PD (1998) *The rise and fall of the plantation complex*. Cambridge University Press. Cambridge.

12. Cavalli-Sforza LL, Bodmer W (1971) *The genetics of human populations*. Freeman. San Francisco.
13. Reed TE (1969) Caucasian genes in American Negroes. *Science* **165**:762-768.
14. Parra EJ, Marcini A, Akey J, Martinson J, Batzer et al. (1998) Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* **63**: 1839–1851.
15. Halder I, Yang B-Z, Kranzler, Stein MB, Shriver MD *et al.* (2009) Measurement of admixture proportions and description of admixture structure in different U.S. populations. *Hum Mut* **30**: 1299-1309.
16. Williams RC, Long JC, Hanson RL, Sievers ML, Knowler (2000) Individual estimates of European genetic admixture associated with lower body-mass index, plasma glucose, and prevalence of type 2 diabetes in Pima Indians. *Am J Hum Genet* **66**: 527-538.
17. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics* **164**: 1567-1587.
18. Tang H, Peng J, Wang P, Risch N (2005) Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol* **28**: 289-301.
19. Ungerer MC, Baird SJE, Pan J, Rieseberg LH (1998) Rapid hybrid speciation in wild sunflowers. *Proc Natl Acad Sci USA* **95**: 11757-11762.
20. Price AL, Tandon A, Paterson N, Barnes KC, Rafaels N *et al.* (2009) Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genetics* **5**(6): e1000519.
21. Pool JE, Nielsen R (2009) Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics* **181**: 711-719.
22. Seldin MF, Morii T, Collins-Schramm HE, Chima B, Kittles R, *et al.* (2004) Putative ancestral origins of chromosomal segments in individual African Americans: implications for admixture mapping. *Genom Res* **14**: 1076-1084.
23. Tang H, Coram M, Wang P, Zhu X, Risch N (2006) Reconstructing genetic ancestry blocks in admixed individuals. *Am J Hum Genet* **79**: 1-12.
24. Schaffner SF (2004) The X chromosome in population genetics. *Nat Rev Genet* **5**: 43-51.
25. Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ *et al.* (2004) A high-density map for disease gene discovery in African Americans. *Am J Hum Genet* **74**: 1001-1013.
26. Adams J, Ward RH (1973) Admixture studies and detection of natural selection. *Science* **180**: 1137-1143.

27. Bryc K, Auton A, Nelson MR, Oksenberg JR, Hauser SL *et al* (2010) Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc Natl Acad Sci USA* **107**: 786-791.
28. Brehm A, Pereira L, Bandelt H-J, Prata MJ, Amorim A (2002) Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Ann Hum Genet* **66**: 49-60.
29. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA *et al.* (2002) A high-resolution recombination map of the human genome. *Nat Genet* **31**: 241-247.
30. Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW *et al.* (1998) Dating the origin of the CCR5- Δ 32 AIDS-resistance allele by the coalescence of haplotypes. *Am J Hum Genet* **62**: 1507-1515.
31. Reich DE, Ruiz Linares A, Goldstein DB (1999) Estimating the age of alleles of mutations using the variation at linked markers. *In* Goldstein DB, Schlötterer C (eds) *Microsatellites: evolution and applications*. Oxford University Press. Oxford.
32. Fenner JN (2005) Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol* **128**: 415-423.